Differential Geometry, Mathematical Physics, and Gauge Theories

An Honors Thesis (HONRS 499)

by

Neal Coleman

Thesis Advisor
Ralph Bremigan

Ralph Bremigan

Ball State University
Muncie, Indiana

May, 2010

Expected Date of Graduation

May 2010

DIFFERENTIAL GEOMETRY, MATHEMATICAL PHYSICS, AND GAUGE THEORIES

NEAL COLEMAN

ABSTRACT. Modern physics relies heavily on geometric modeling of reality. It can be effectively phrased in terms of differential forms, vector fields, and other objects of differential geometry, and manipulated with the tools of differential geometry. Differential geometry itself relies on the development of advanced linear algebra, including the theory of tensor spaces and forms over vector spaces. Therefore, I develop advanced linear algebra, aspects of manifold theory, elements of pseudo-Riemannian geometry, a brief introduction to Lie groups, and some statements regarding vector bundles, in order to convey some aspects of modern physical theories, including the classical theory of electrodynamics, the theories of special and general relativity, and some notions of gauge theories.

ACKNOWLEDGMENTS

- I wish to thank my advisor, Dr. Ralph Bremigan, for patience, encouragement, and mentorship not only during this project but also through my entire mathematical journey so far.
- I also would like to thank Dr. Tony Cancio, for hours of invaluable discussion regarding gauge theories and electrodynamics.

CONTENTS

Foreword

As far as I'm concerned, none of the material you're about to read is original. However, I have done my best to general digest[1] it, restate it comprehensibly, and link different portions

[1]Much like the worms a mother bird feeds her hatchlings.

together; to my knowledge, the manner in which I have structured the theory — building it beginning with advanced linear algebra and working through to bundles — is original, although Göckeler and Schücker ([1]) and Warner ([3]) follow similar but more limited approaches in their works. I have tussled with the material and rewritten it in an attempt to better understand it myself, and hopefully to present it in an enlightening manner to an interested audience. When appropriate, I have tried to indicate where the interested reader can find the original work. At the end, I have tried to list all of the sources I have consulted in preparing this thesis, as well as some I have not, so you can track down authoritative material.

Some words to the most important: thank you to my wife, Rachel, and daughter, Claire, for your infinite patience with me.

Introduction

Methodological naturalism is the description of nature through repeated revision of models based on comparison of the models' predictions to observation. In precise sciences — especially physics — the models take the form of a set of mathematical axioms from which deductions are made in order to predict particular patterns of physical behavior. These patterns are then compared to the patterns actually observed in nature, and the axioms are confirmed, revised, or discarded accordingly.

It is amazing how much physics can be phrased in terms of geometric mathematics — or, equivalently, how much of nature's fundamental patterns can be explained by thinking geometrically. The classical example is the theory of general relativity, upon which we shall touch in due course, but many other hypotheses can be thought of geometrically, from electromagnetism to the fundamental Standard Model of particle physics.

After spending the bulk of our time elaborating on the mathematical theory, we shall explore the physics of these models, especially particle physics, as we see how differential geometry serves as a language for physical expression.

Part 1. The Mathematics

1. Linear Algebra

1.1. Motivation.

Advanced linear algebra, the theory of vector spaces, is necessary to any study of differential geometry. Differential geometry is the study of "differentiable manifolds," spaces with differentiable structures which are described in terms of vector spaces. One might regard a differentiable manifold as many vector spaces sewn together. Therefore, it is advantageous to review linear algebra.

The following section largely follows Chapter 2 of [2], although it is presented in terms of general vector spaces.

1.2. Preliminaries.

In all that follows, we assume familiarity with elementary linear algebra.

A *module* K over a ring R is an abelian group K and an operation *scalar multiplication* $R \times K \to K$ such that for all $r, s \in R$ and $v, w \in K$

- $r(v + w) = rv + rw$,
- $(r + s)v = rv + rs$,
- $(rs)v = r(sv)$, and

- $1_R v = v$ if R has an identity 1_R.

A *real vector space* is a module over the field of the real numbers.

Let V be a finite-dimensional vector space. We define the *dual space* of V to be $V^* := \{\omega : V \to \mathbb{R} \mid \omega \text{ is linear}\}$. Let $E = \{e_1, e_2, \ldots, e_n\}$ be any basis of V. Define the *dual basis* $E' = \{e^1, e^2, \ldots, e^n\}$ of V^*, where e^i is the function such that $e^i(e_j) = \delta_{ij}$. If we let $\omega \in V^*$, then since any $v \in V$ can be written as $v = \sum v_i e_i$, we have $\omega(v) = \sum v_i \omega(e_i)$. But $(\sum \omega(e_i)e^i)(v) = (\sum \omega(e_i)e^i)(\sum v_j e_j) = \sum_i \sum_j v_j \omega(e_i)\delta_{ij} = \sum v_i \omega(e_i)$, so we conclude $\omega = \sum \omega(e_i)e^i$ and so E' spans V^*. To see that E' is linearly independent, consider the zero $0 \in V^*$ as a linear combination of elements of E', $0 = \sum k_i e^i$. Because $0(v) = 0$ for each $v \in V$, $\sum k_i e^i(e_i) = k_i = 0(e_i) = 0$ for each i, so E' is linearly independent and is therefore a basis of V^*.

The dual of the dual, V^{**}, is canonically isomorphic to V when V has finite dimension.[2] For let $f : V \to V^{**}$ be defined by $f(v)(\omega) = \omega(v)$ for each $\omega \in V^*$. It is clear that f is a homomorphism of vector spaces by the linearity of elements of V^*: $f(rv + w)(\omega) = \omega(rv + w) = r\omega(v) + \omega(w) = rf(v)(\omega) + f(w)(\omega)$ for each $\omega \in V^*$. Let $S \in V^{**}$; then, for $s = \sum S(e^i)e_i \in V$, we have $f(s) = S$ since $S(\omega) = \sum S(e^i)\omega(e_i)$ and $f(s)(\omega) = \sum S(e^i)f(e_i)(\omega) = \sum S(e^i)\omega(e_i)$ for all $\omega \in V^*$. Thus, f is surjective. Moreover, if $f(v) = f(w)$ for $v, w \in V$, then for each $\omega \in V^*$, we have that $f(v)(\omega) = f(w)(\omega)$ implies $\omega(v) = \omega(w)$. Writing in terms of a basis of V, $\omega(\sum v_i e_i) = \omega(\sum w_i e_i)$ gives $\sum v_i \omega(e_i) = \sum w_i \omega(e_i)$, so $v_i = w_i$. Thus $v = w$, and so f is injective. Therefore, f, being a bijective homomorphism, is an isomorphism of vector spaces.

1.3. **Tensors.** The set

$$\mathcal{T}_s^r = \{A : (V^*)^r \times V^s \to \mathbb{R} \mid A \text{ is } \mathbb{R}\text{-multilinear in each argument}\}$$

is called the *tensors of type* (r, s) *over* V.[3] By convention, we set $\mathcal{T}_0^0 = \mathbb{R}$. We assume that V is apparent from context.

Tensors are not just of intrinsic interest in the theory of vector spaces, but also of great use in describing *basis-independent relationships* among vectors. Many of the most important results of differential geometry are described in terms of tensors — for example, the Riemann curvature tensor encodes the curvature of a surface as a relationship between four vector fields in a manner independent of coordinates, which is a generalization of Gauss' *Theorema Egregium* that curvature is an intrinsic property of a surface. Tensors are also useful in modern physics as a consequence of the *Galilean Principle*: the laws of nature are identical in *all* inertial reference frames. Since natural laws are modeled in terms of relationships among vector fields, tensor fields are natural[4] ways to describe them.

While we are discussing physics, a note here on the difference between physicists' and mathematicians' conceptions of tensors. Physicists, being interested in the covariance of natural laws, tend to formulate the the mathematics of tensors in terms of coordinates without much thought as to what tensors actually *are*. This pragmatic *notio rerum* can lead to to mathematically imprecise definitions of tensors.

[2]The two are isomorphic when V has infinite dimension as well, but there is no canonical isomorphism.

[3]I follow this definition because I find it most intuitive. It is possible to define tensors in a different, and more general manner (c.f., [3], pp. 54 – 5).

[4]Pun unintended.

Let us now develop some properties of tensors over a vector space V. Note that $\mathcal{T}_1^0 = V^*$ and $\mathcal{T}_0^1 = V$ by the duality mentioned above. Also, if $A : V^s \to V$, put $\bar{A} \in \mathcal{T}_s^1$ to be $\bar{A}(\omega, v_1, \ldots, v_s) = \omega(A(v_1, \ldots, v_s))$ for each $\omega \in V^*$ and $v_i \in V$, $i = 1, \ldots, s$.

If any of the arguments of a type (r, s) tensor are zero, then the tensor itself evaluates to zero. For without loss of generality[5] let $v_i = 0$; then, in terms of a basis e_j on V, $v_i = \sum v_j e_j$ and we have all the $v_j = 0$. Then since tensors are multilinear we have, for any $A \in \mathcal{T}_s^r$, $A(\omega_1, \ldots, v_i, \ldots, v_s) = A(\omega_1, \ldots, \sum v_j e_j, \ldots, v_s) = \sum v_j A(\omega_1, \ldots, e_j, \ldots, v_s) = 0$.

Tensors of type (r, s) form a vector space over $\mathbb{R}$. It is possible to add additional structure by defining *tensor multiplication* or the *tensor product*

$$\otimes : \mathcal{T}_q^p \times \mathcal{T}_s^r \to \mathcal{T}_{q+s}^{p+r}$$

as follows: if $A \in \mathcal{T}_q^p$ and $B \in \mathcal{T}_s^r$, then

$$\begin{aligned} &(A \otimes B)(\omega_1, \ldots, \omega_{p+r}, v_1 \ldots, v_{q+s}) \\ &= A(\omega_1, \ldots, \omega_p, v_1, \ldots, v_q) B(\omega_{p+1}, \ldots, \omega_{p+r}, v_{q+1}, \ldots, v_{q+s}). \end{aligned}$$

(Multiplication is permitted because tensors map sets of vectors and one-forms to $\mathbb{R}$.) In addition to being well-defined, the map is also bilinear, and so obeys the distributive law: $((rA + sA') \otimes B)(\cdot, \cdot) = (rA + sA')(\cdot)B(\cdot) = (rA(\cdot) + sA'(\cdot))B(\cdot) = rA(\cdot)B(\cdot) + sA'(\cdot)B(\cdot) = rA \otimes B + sA' \otimes B$. Moreover, the tensor product is associative since the real numbers are associative. However, it is *not* in general commutative; for example, take a two-dimensional vector space with basis $\{e_1, e_2\}$ and its dual space with the dual basis $\{e^1, e^2\}$. Then $e^1 \otimes e^2 \in \mathcal{T}_2^0$. We have $(e^1 \otimes e^2)(e_1, e_2) = \delta_{11}\delta_{22} = 1$ but $(e^2 \otimes e^1)(e_1, e_2) = \delta_{12}\delta_{21} = 0$.

We call tensors of type $(0, s)$ *covariant* and tensors of type $(r, 0)$ *contravariant.* Note that if $A \in \mathcal{T}_s^0$ and $B \in \mathcal{T}_0^r$, $A \otimes B = B \otimes A$.

Much like vectors and forms, tensors are defined independently of basis. However, it is often convenient to work with tensors in bases.[6] If we give V (of dimension n) a basis $\{e_i\}$ and V^* the dual basis $\{e^i\}$, then, for a tensor $A \in \mathcal{T}_s^r$, the *components of A relative to the basis* $\{e_i\}$ are the real numbers $A_{j_1 \ldots j_s}^{i_1 \ldots i_r} = A(e^{i_1}, \ldots, e^{i_r}, e_{j_1}, \ldots, e_{j_s})$, where $i_k, j_l = 1, \ldots, n$ for $k = 1, \ldots, r$ and $l = 1, \ldots, s$. As a sanity check, a $(0, 1)$ tensor is an element $\omega \in V^*$, and according to this definition, it has components $\omega(e_i)$, which are exactly the components used with the dual basis: $\omega = \sum \omega(e_i)e^i$.

It is a little more difficult to check that this is true for vectors; remember that we have the canonical isomorphism $f : V \to V^{**} = \mathcal{T}_0^1$ defined above by $f(v)(\omega) = \omega(v)$ for each $\omega \in V^*$. Since $f(e_i)$ maps a form to its i^{th} component, f induces the basis $\{f(e_i)\}$ corresponding to $\{e_i\}$. Hence the components of a tensor $A \in \mathcal{T}_0^1$ are $A^i = A(e^i) = e^i(A)$, which is the i^{th} component of A when interpreted as a vector.

Components are also useful when evaluating tensors on vectors and forms. Following [2], p. 39, take, as an example, a type $(1, 2)$ tensor. If we write the two vectors v, w and one form ω out in terms of components, we have $\omega = \sum \omega^k e^k$, $v = \sum v_i e_i$, and $w = \sum w_j e_j$. Then, using multilinearity, $A(\omega, v, w) = A(\sum \omega^k e^k, \sum v_i e_i, \sum w_j e_j) = \sum_{i,j,k} \omega^k v_i w_j A(e^k, e_i, e_j) = \sum_{i,j,k} A_{ij}^k v_i w_j \omega^k$. The general case is similar, and only requires keeping careful track of dots, superscripts, and subscripts.

---

[5]Use the dual basis $\{e^j\}$ of V^* if you want to do the proof with the dual space.

[6]As I mentioned above, this is what physicists like to do.

If $A \in \mathcal{T}^p_q$ and $B \in \mathcal{T}^r_s$, the components of $A \otimes B$ are given by

$$(A \otimes B)^{i_1 \dots i_{p+r}}_{j_1 \dots j_{q+s}} = A^{i_1 \dots i_p}_{j_1 \dots j_q} B^{i_{p+1} \dots i_{p+s}}_{j_{q+1} \dots j_{q+s}}.$$

This follows directly from the components of tensor components and multiplication.

Finally, the component notation permits us to find[7] a basis of $\mathcal{T}^r_s$ induced by $\{e_i\}$. Namely, given a tensor $A \in \mathcal{T}^r_s$, $A = \sum A^{i_1 \dots i_r}_{j_1 \dots j_s} e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_r} \otimes e^{j_1} \otimes \dots \otimes e^{j_s}$ because $(e_{i_1} \otimes e_{i_2} \otimes \dots \otimes e_{i_r} \otimes e^{j_1} \otimes \dots \otimes e^{j_s})(e^{k_1}, \dots, e^{k_r}, e_{l_1}, \dots, e_{l_s}) = \delta^{k_1}_{i_1} \delta^{k_2}_{i_2} \dots \delta^{k_r}_{i_r} \delta^{j_1}_{l_1} \dots \delta^{j_s}_{l_s}$.

Tensors transform from basis to basis in a consistent manner, which we will find useful in Section 2.3. Let $\{\epsilon_i\}$ and $\{e_i\}$ be bases on V which induce dual bases $\{\epsilon^i\}$ and $\{e^i\}$ respectively. If $L = \lambda_{ij}$ is the linear transformation (change-of-basis matrix) relating $\{\epsilon_i\}$ and $\{e_i\}$, and $L^* = \lambda^{ij}$ is the associated change of basis transformation on V^*, then the tensor $A \in \mathcal{T}^r_s$ with components $A^{i_1 \dots i_r}_{j_1 \dots j_s}$ under $\{\epsilon_i\}$ and $\{\epsilon^i\}$ has components $A^{k_1 \dots k_r}_{l_1 \dots l_s}$ under $\{e_i\}$ and $\{e^i\}$; the components are related by

$$A^{k_1 \dots k_r}_{l_1 \dots l_s} = \sum \lambda^{k_1 i_1} \lambda^{k_2 i_2} \dots \lambda^{k_r i_r} \lambda_{l_1 j_1} \cdots \lambda_{l_s j_s} A^{i_1 \dots i_r}_{j_1 \dots j_s}.$$

For example, if $A \in \mathcal{T}^1_2$, $A^k_{ij} = A(e^k, e_i, e_j)$ and $A^l_{mn} = A(\epsilon^l, \epsilon_m, \epsilon_n)$. Since $e^k = \sum_l \lambda^{kl} \epsilon^l$, $e_i = \sum_m \lambda_{im} \epsilon_m$, and $e_j = \sum_n \lambda_{jn} \epsilon_n$, we have

$$\begin{aligned}
A^k_{ij} &= A(e^k, e_i, e_j) \\
&= A(\sum_l \lambda^{kl} \epsilon^l, \sum_m \lambda_{im} \epsilon_m, \sum_n \lambda_{jn} \epsilon_n) \\
&= \sum_{l,m,n} \lambda^{kl} \lambda_{im} \lambda_{jn} A(\epsilon^l, \epsilon_m, \epsilon_n) \\
&= \sum_{l,m,n} \lambda^{kl} \lambda_{im} \lambda_{jn} A^l_{mn}.
\end{aligned}$$

The operation of *contraction* is a generalization of the trace function on matrices (which are representations of $(1,1)$, $(0,2)$, and $(2,0)$ tensors). The $(1,1)$ contraction operation is the unique linear function $C : \mathcal{T}^1_1 \to \mathbb{R}$ such that $C(v \otimes \omega) = \omega(v)$ for all $v \in V$ and $\omega \in V^*$. Given any basis $\{e_i\}$ of V, we can write $A = \sum A^i_j e_i \otimes e^j$. $C(e_i \otimes e^j) = e^j(e_i) = \delta^j_i$, so we define $C(A) = \sum A^i_i = \sum A(e^i, e_j)$. We can easily extend this to a function $C^i_j : \mathcal{T}^r_s \to \mathcal{T}^{r-1}_{s-1}$. *Fix* $\omega_1, \dots, \omega_{r-s}$ and $v_1, \dots, v_{s-1}$ and define $\bar{A} \in \mathcal{T}^1_1$ by $\bar{A} = A(\omega_1, \dots, \omega_{i-1}, \omega, \omega_i, \dots, \omega_{r-s}, v_1, \dots, v_{j-1}, v, v_j, \dots, v_s)$. Now we define $C^i_j A = C\bar{A}$. In components, this amounts to setting the i^{th} contravariant index equal to the j^{th} covariant index and summing: if A has components $A^{i_1 \dots i_r}_{j_1 \dots j_s}$, $C^i_j A$ has components $\sum_m A^{i_1 \dots i_{i-1} m i_{i+1} \dots i_r}_{j_1 \dots j_{j-1} m j_{j+1} \dots j_s}$. More information can be found in [2], pp. 40 – 41.

1.4. **Forms.** A subspace of the tensors will especially concern us.[8] The *differential k-forms of V* are $\Lambda^k = \{A \in \mathcal{T}^0_k | A \text{ is skew-symmetric}\}$[9]. That is, if $\omega \in \Lambda^p$, then

$$\omega(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -\omega(v_1, \dots, v_j, \dots, v_i, \dots, v_n)$$

[7]Parsing this boils down to matching indices and using the tensor multiplication rules.

[8][3] develops this differently, as he develops tensor spaces differently — c.f. p. 56.

[9]"Alternating" is another word for "skew-symmetric"

for each $i, j = 1, \ldots, n, i \neq j$. Note that Λ^m is trivial if $m > n$ since elements of Λ^p send linearly dependent sets of p vectors to 0 in $\mathbb{R}$.

We define a function $\wedge : \Lambda^p \times \Lambda^q \to \Lambda^{p+q}$ by, if $\psi \in \Lambda^p$ and $\omega \in \Lambda^q$,

$$(\psi \wedge \omega)(v_1, \ldots, v_{p+1}) = \frac{1}{p!q!} \sum_{\pi \in S_p} \psi(v_{\pi(1)}, \ldots, v_{\pi(p)})\omega(v_{\pi(p+1)}, \ldots, v_{\pi(p_q)})sgn(\pi),$$

where S_p is the symmetric group of bijections from a set of p elements to itself.

1.5. **Bilinear Forms.** Tensors of type $(0, 2)$ are *bilinear forms.* A bilinear form $b \in \mathcal{T}_2^0$ is *symmetric* if $b(v, w) = b(w, v)$ for each $v, w \in V$. There are five interesting kinds of symmetric bilinear forms: *positive [negative] definite* ($v \neq 0 \Rightarrow b(v, v) > 0[b(v, v) < 0]$); *positive [negative] semidefinite* ($v \neq 0 \Rightarrow b(v, v) \geq 0[b(v, v) \leq 0]$); and *nondegenerate* ($b(v, w) = 0$ for all $w \in V$ implies $v = 0$). If b is symmetric and bilinear, then b restricted to any subspace W in V is again a symmetric bilinear form; the same holds for (semi-)definiteness.

Symmetric bilinear forms are useful because they generalize inner products. More information on this generalization can be found in [2], pp. 46 – 52.

2. Differentiable Manifolds

A differentiable manifold, briefly, is a space with a set of overlapping maps from the space to $\mathbb{R}^n$ which permit differentiation to take place "on" the manifold by doing calculus in $\mathbb{R}^n$ by proxy. This ability to differentiate permits us to do differential and integral calculus and leads to generalizations of Stokes' Theorem, possibly one of the most important results in the field. The ability to work in $\mathbb{R}^n$ by proxy also endows the space with a vector space structures, which permits us to apply linear algebra.

A number of texts develop this theory, including [2], [3], [1], [4], [5]. [4] is regarded as reference-grade.

2.1. **Preliminaries.** Let M be a topological space. M is said to be *locally Euclidean* if for each $p \in M$ there is an open neighborhood U of p with a homeomorphism $\xi : U \to \mathbb{R}^n$. (To avoid pinch points, edges, and the like, we require that every point in M have an open neighborhood homeomorphic to an open set of $\mathbb{R}^n$, and we say that M has *dimension n.*) The map ξ is called a *coordinate map.* We can compose ξ with the Euclidean projection functions $\pi_i : \mathbb{R}^n \to \mathbb{R}$, which send $p \to p_i$, the i^{th} coordinate of p, to get the *coordinate functions* $x^i = \pi_i \circ \xi$. If two coordinate maps ξ and η are both defined on an open set $U \subseteq M$, then we say the overlap is *smooth* if the maps $\xi \circ \eta^{-1} : \eta(U) \to \xi(U)$ and $\eta \circ \xi^{-1} : \xi(U) \to \eta(U)$ are C^∞ functions.[10] (The condition holds trivially if the domains of η and ξ do not meet.) A *coordinate system* is a coordinate map and domain.

We now deal with collections of smoothly overlapping functions. Given a locally Euclidean space M, an *atlas* on M is a set $\mathcal{A}$ of smoothly overlapping coordinate maps such that their domains cover M. In the tradition of Borel sets, complete topologies, and other maximal collections, a *differentiable structure, maximal atlas,* or *complete atlas* $\mathcal{A}$ on M is an atlas such that the following condition holds: if (U, ξ) is a coordinate system that overlaps smoothly with every coordinate system in $\mathcal{A}$, then $(U, \xi) \in \mathcal{A}$. It just so happens that *any* atlas $\mathcal{A}$ on M is contained in a unique differentiable structure, the set of all coordinate systems

[10]It is certainly possible to consider C^k functions, for some $0 \leq k < \infty$, but I shall not do so here. The theory is similar, but not identical. For some development, see [3].

that overlap with coordinate systems in $\mathcal{A}$ (which is again a coordinate system for obvious reasons).

A *smooth manifold* (often just *manifold*) M is locally Euclidean space furnished with a differentiable structure.

2.2. **Mappings.** Let M be a manifold of dimension m. A function $f : M \to \mathbb{R}$ is said to be *smooth* if, for any coordinate system (U, ξ) in M, on U the function $f \circ \xi^{-1} : \xi(U) \to \mathbb{R}$ is C^∞. The set of all such smooth functions on M is denoted $\mathcal{F}(M)$. If N is a manifold of dimension n and $f : M \to N$, then, as before, we define smoothness by using the coordinate systems to drop back into familiar territory: f is smooth provided that, for any coordinate system (U, ξ) on M and (V, η) on N, when f is restricted to the domains of the coordinate systems, the function $\eta \circ f \circ \xi^{-1} : \xi(U) \to \eta(V)$ is a C^∞ map between open sets of $\mathbb{R}^{\dim(M)}$ and $\mathbb{R}^{\dim(N)}$. In this case, we write $f \in \mathcal{F}(M, N)$.

The composition of smooth maps is again smooth because coordinate systems are required to overlap smoothly. Coordinate systems are themselves smooth maps from open sets of a manifold to Euclidean space. Smoothness is *local* – it is defined on neighborhoods and only incidentally becomes global upon our insistence that the entire manifold be involved (it does not tax the imagination to create functions that are only smooth in some neighborhoods in a manifold).

To motivate the discussion, let us compare morphisms in set theory, algebra, and topology. In set theory, we have simple counting maps, injections and surjections. If a map is both injective and surjective, it is said to be a bijection; up to cardinality, it is impossible to distinguish sets related by a bijection. In group theory, homomorphisms are maps between groups which preserve the group operation. If a map between two groups is a homomorphism and its inverse is also a homomorphism, the map is said to be an isomorphism, and we need not distinguish between the algebraic structures — they are equivalent. In topology, continuous functions permit us to compare open sets; if two topological spaces have between them a continuous function with a continuous inverse, the function is known as a homeomorphism, and the topologies of the spaces are considered equivalent.[11]

We are therefore led to consider maps which preserve the differential structure. A smooth map between two manifolds with a smooth inverse is called a *diffeomorphism*, and if two manifolds have such a map between them, they are called *diffeomorphic*, and we may regard their differential structures as equivalent. Of course, a diffeomorphism induces a homeomorphism: since the topology of M is related to the choice of differentiable structure, preservation of differentiable structure implies preservation of topological structure. However, a homeomorphism between two manifolds is not enough to conclude that they are diffeomorphic.[12]

2.3. **Tangent vectors, tensors, and forms.** In mathematics, when we wish to study the *abstract* behavior of relatively concrete objects, we axiomatize their essential properties and generalize the theory. This is the case with tangent vectors. We are motivated by the study of two-dimensional surfaces embedded in $\mathbb{R}^3$, which possess literal tangent planes, described by the gradient of a local coordinate patch. For example, if a two-dimensional surface Γ in

[11]I wish this comparison were mine, but it is not. O'Neill ([2]) lays it out nicely in Table 1 on p. 93.

[12]O'Neill points out $t \mapsto t^3$ between the real line and itself. There are also non-manifolds that are homeomorphic to manifolds: for example, a cube and a sphere.

$\mathbb{R}^3$ is parameterized by $f : \mathbb{R}^2 \to \mathbb{R}^3$ given by $f(u, v) = (x(u, v), y(u, v), z(u, v))$, at $f(u_0, v_0)$ the plane spanned by the vectors $\frac{\partial f}{\partial u}|_{(u_0,v_0)}$ and $\frac{\partial f}{\partial v}|_{(u_0,v_0)}$ is the tangent plane.

Following this, we motivate the abstraction of tangent vectors by noting that the intrinsic property of a tangent vector is that it defines a *directional derivative*: on a surface a function can be differentiated at a point with respect to a direction.[13]

Given a smooth manifold M, fix a point $p \in M$. A *tangent vector to* M *at* p[14] or *linear derivation of* $\mathcal{F}(M)$ *at* p is a function $v : \mathcal{F}(M) \to \mathbb{R}$ such that, for all $a, b \in \mathbb{R}$, $f, g \in \mathcal{F}$,

(1) $v(af + bg) = av(f) + bv(g)$ and
(2) $v(fg) = gv(f) + fv(g)$.

The *tangent space to* M *at* p, denoted T_pM, is the collection of all tangent vectors to M at p. It can be helpful to think of T_pM as a copy of $\mathbb{R}^n$ attached to M at p. The set T_pM is a vector space if we define, for $v, w \in T_pM$, $r \in \mathbb{R}$, and $f \in \mathcal{F}$, $(v + w)(f) = v(f) + w(f)$ and $(rv)(f) = r(v(f))$.

Because T_pM is a vector space, we can apply the linear algebra we previously developed. If ξ is a coordinate system in M at p and $f \in \mathcal{F}$, put

$$\frac{\partial f}{\partial x^i}|_p = \frac{\partial(f \circ \xi^{-1})}{\partial \pi_i}|\xi(p).$$

It turns out that $\partial_i = \partial x_i \frac{\partial}{\partial x^i} : \mathcal{F} \to \mathbb{R}$ is a tangent vector; in fact, the set of partial derivatives $\{\partial_i\}$ is a basis on T_pM. If we have two coordinate systems $\phi = \{x_i\}, \psi = \{y_i\} : U \to \mathbb{R}^n$, we have two sets of bases, respectively $\{\frac{\partial}{\partial x^i}\}$ and $\{\frac{\partial}{\partial y^i}\}$. If $f \in \mathcal{F}(M)$ is chosen arbitrarily, then we can formulate a change-of-basis matrix with the chain rule:

$$\frac{\partial f}{\partial x^i} = \sum_j \frac{\partial y^j}{\partial x^i}\frac{\partial f}{\partial y^j}.$$

That is, written without reference to the arbitrary function,

$$\partial x_i = \frac{\partial}{\partial x^i} = \sum_j \frac{\partial y^j}{\partial x^i}\frac{\partial}{\partial y^j}.$$

We now *automatically* have tensor spaces at each point. By picking a coordinate system, we may induces bases on the related tensor spaces. We restate the characterization of dual and tensor space bases from the previous section because notation has changed. The *cotangent space to* M *at* p, T_pM^*, is the dual space to T_pM, and has a dual basis $\{dx^i\}$, where the dx^i are characterized by $dx^i(\partial_j) = \delta^i_j$. The tensors of type (r, s) over T_pM carry an induced basis $\partial_{i_1} \otimes \ldots \otimes \partial_{i_r} \otimes dx^{j_1} \otimes dx^{j_s}$. The k-forms at p, Λ^k over T_pM, carry an induced basis $dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_k}$. (The space k-forms comprise the trivial ring if $k > n$, and the dimension of Λ^k over T_pM is $\binom{n}{k}$; the dimension of $\mathcal{T}^r_s$ is n^{rs}.)

In terms of a basis, to repeat Section 1.4 above in the present notation, a tensor A of type (r, s) has rs components $A^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ defined by

$$A^{i_1 \ldots i_r}_{j_1 \ldots j_s} = A(dx^{i_1}, \ldots, dx^{i_r}, \partial_{j_1}, \ldots, \partial_{j_s}).$$

[13]In all that follows, bear in mind that tangent vectors, at heart, are really just arrows sticking off a surface.

[14]Warner is characteristically more technical; see [3], pp. 11 – 14.

The components of tensors and forms transform according to the rules laid out in section 3.3. If $\{x_i\}$ and $\{y_i\}$ are two sets of coordinate functions, to determine the transformation rule it is enough to determine how basis elements transform. Let $\partial_{i_1} \otimes \ldots \otimes \partial_{i_r} \otimes dx^{j_1} \otimes \ldots \otimes dx^{j_s}$ be a basis element of $\mathcal{R}^r_s$. In $\{y_i\}$, this tensor is a linear combination of the basis elements induced by the coordinate functions:

$$\partial x_{i_1} \otimes \ldots \otimes \partial x_{i_r} \otimes dx^{j_1} \otimes \ldots \otimes dx^{j_s} = \sum A^{k_1 \ldots k_r}_{l_1 \ldots l_s} \partial y_{k_1} \otimes \ldots \otimes \partial y_{k_r} \otimes dy^{l_1} \otimes \ldots \otimes dy^{l_s}.$$

The components can be found by checking how the tensor operates on basis elements:

$$A^{k_1 \ldots k_r}_{l_1 \ldots l_s} = (\partial x_{i_1} \otimes \ldots \otimes \partial x_{i_r} \otimes dx^{j_1} \otimes \ldots \otimes dx^{j_s})(dy^{i_1}, \ldots, dy^{i_r}, \partial y_{j_1}, \ldots, \partial y_{j_s}).$$

Interpreting $\partial_i(dx^j) = dx^j(\partial_i)$, we have

$$\begin{aligned} A^{k_1 \ldots k_r}_{l_1 \ldots l_s} &= \partial x_{i_1}(dy^{k_1}) \cdots \partial x_{i_r}(dy^{k_r}) dx^{j_1}(\partial y_{l_1} \cdots dx^{j_s}(\partial y_{l_s}) \\ &= \frac{\partial y_{k_1}}{\partial x_{i_1}} \cdots \frac{\partial y_{k_r}}{\partial x_{i_r}} \frac{\partial x_{j_1}}{\partial y_{l_1}} \cdots \frac{\partial x_{j_s}}{\partial y_{l_s}}, \end{aligned}$$

since the one-forms pick out a single term from each coordinate-change transformation.

We wish to ultimately do calculus on M, so we concern ourselves not just with the vector spaces alone, but how they relate to each other. We want to deal with the *local* characteristics of maps from the manifold to the vector spaces, not just the *pointwise* characteristics. Consequently, we need to have some notion of "smoothness." At this point, there is no relationship between T_pM and T_qM for any $p, q \in M$; the two are entirely independent unless $p = q$.

A *vector field*[15] (*tensor field*) (*k-form*) on a manifold M is a choice of vector (tensor) (k-form) at each point. We can consider a vector field V on M as a function on $\mathcal{F}(M)$ which assigns to $f \in \mathcal{F}$ some function Vf. We say V is *smooth* if Vf is smooth for all $f \in \mathcal{F}(M)$, i.e., if $V : \mathcal{F}(M) \to \mathcal{F}(M)$. We denote the set of all smooth vector fields by $\mathfrak{X}(M)$. Similarly, smooth one-forms (elements of $T^*(M)$) can be regarded as maps from $\mathfrak{X}(M)$ to $\mathcal{F}(M)$. A tensor field $A \in \mathcal{T}^r_s(M)$ is smooth if $A : (\mathfrak{X}^*(M))^r \times \mathfrak{X}(M)^s \to \mathcal{F}(M)$.

The operation $[\cdot, \cdot] : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ on vector spaces defined by $([V, W])f = V(Wf) - W(Vf)$ is termed the *Lie bracket*. It is anticommutative, $\mathbb{R}$-multilinear, and for each $f, g \in \mathcal{F}(M)$ and $U, V, W \in \mathfrak{X}(M)$, has $[[U, V], W] + [[V, W], U] + [[W, U], V] = 0$ (the *Jacobi identity*) and $[fV, gW] = fg[V, W] + f(Vg)W - g(Wf)V$.

2.4. **Tensor calculus.** The first step in constructing a general calculus of tensors is to revisit differentiation of functions on manifolds. If M, N are manifolds and $f : M \to N$ is a C^∞ function, we may define the *differential of f df* to be a smoothly varying pointwise transformation of *tangent spaces*: for each $p \in M$, $df : T_pM \to T_{f(p)}N$ is given by $(df(v))(g) = v(g \circ f)$. That is, the tangent vector $df(v) \in T_{f(p)}N$, operating on an arbitrary function $g : N \to \mathbb{R}$, maps to exactly the same number as $v \in T_pM$ does acting on the function $g \circ f : M \to \mathbb{R}$.

Writing this down in local coordinates, the differential of f is entirely determined by its action on the basis elements of T_pM. If $\phi = \{x_1, \ldots, x_n\}$ is a coordinate system at $p \in M$

[15]Here, we follow O'Neill's definitions; Warner's ([3]) are different, and involve constructing a differential structure on the appropriate bundles.

and $\psi = \{y_1, \ldots, y_m\}$ a coordinate system at $f(p) \in N$, we recall that a vector $w \in T_{f(p)}M$ is characterized in the $\{y_j\}$ by $w = \sum_j w(y_j)\partial_j$. Immediately,

$$df\left(\frac{\partial}{\partial x_i}\right) = \sum_j \left[df\left(\frac{\partial}{\partial x_i}\right)(y_j)\right]\frac{\partial}{\partial y_j} = \sum_j \frac{\partial(y_j \circ f)}{\partial x_i}\frac{\partial}{\partial y_j}.$$

The matrix

$$J_{ij} = \left(\frac{\partial(y_j \circ f)}{\partial x_i}\right)$$

relative to ϕ and ψ is called the *Jacobian of* f.

Note in particular that if we have a curve $\sigma : (0,1) \to M$, we can use the fact that the interval $(0,1)$ is a manifold in order to differentiate σ. The time-derivative of a point on $\sigma(0,1)$ corresponds to the motion of the image of a point in $(0,1)$ as it moves along the interval at constant speed. Therefore, we define

$$\sigma'(t) = d\sigma\left(\frac{d}{dt}|_t\right).$$

In the coordinate system $\{x_1, \ldots, x_n\}$, we have $\sigma'(t) = \sum \frac{d(x_i \circ \sigma)}{dt}\partial_i$.

We can define two more maps induced by d. Let $\delta : T_{f(p)}N^* \to T_pM^*$ be given by $\delta f(\theta)(v) = \theta(df(v))$ for $v \in T_pM, \theta \in T_{f(p)}M^*$ and, *if df is invertible at p*, lat $\partial : T_pM^* \to T_{f(p)}N^*$ be given by $\partial f(\theta)(v) = v(df^{-1}v)$ for $\theta \in T_pM^*$ and $v \in T_{f(p)}N^*$.

We would like to generalize the notion of directional derivative.[16] In order to do this, we need to be able to relate tangent vectors with directions on M. If X is a vector field, it should be possible to "flow" along X, like a leaf floating along a stream. So, for an interval $I \subseteq \mathbb{R}$, we define an *integral curve* of X as a smooth curve $\sigma : I \to M$ such that, at each $t \in I$, $\sigma'(t) = X(\sigma_X(t))$. Fixing $X \in \mathfrak{X}(M)$, from the uniqueness and existence theorem of first-order ordinary differential equations,[17] for each $p \in M$ there is a unique integral curve σ_p defined on a domain (a_p, b_p) such that $\sigma_p(0) = p$. The endpoints of the domain of the curve in general depend on p.

Let $W = \{(t,p) \in \mathbb{R} \times M \mid : a_p < t < b_p\}$, and define $\Phi : W \to M$ by $\Phi(t,p) = \sigma_p(t)$. Interpret it this way: $\Phi(t,p)$ is the point where a leaf is after time t if we put it in the water at p. Now set $M_t = \{p \in M \mid a_p < t < b_p\}$. This is the projection of W with a fixed t. The set M_t is open in M. Finally, define the transformation $\phi_t : M_t \to M$ by $\phi_t(p) = \Phi(t,p)$. This is the "flow" of M through X up to a time t. If the flow would have to stop, say at a hole in the manifold, M is pared down to M_t.

The ϕ_ts have several nice properties: $M_{-t} = \phi_t(M_t)$ and, on M_{s+t}, $\phi_s \circ \phi_t = \phi_{s+t}$. The ϕ_ts are called the *one-parameter group of local transformations of* M. A vector field on M is called *complete* if at each point of M its integral curve is defined over all of $\mathbb{R}$. Equivalently, X is complete if the domain of each of the elements of its one-parameter group of local transformations is all of M.

We are now in a position to consider the flow of a tensor field relative to a vector field. Because the ϕ_t are diffeomorphisms, $d\phi_t$ is invertible. Consequently, we are able to "pull

[16]Here, we follow Morita ([6]), pp. 39-43. Warner ([3]) has a similar exposition (pp. 37-40.

[17]See any book on ordinary differential equations.

back" tensors from a short ϵ-distance away to p. More generally, for an invertible function $f: M \to N$ acting on $\mathcal{T}^r_s(N)$, we define

$$(f^*A)(\theta_1,\ldots,\theta_r,v_1,\ldots,v_s) = A(\partial f(\theta_1),\ldots,\partial f(\theta_r),df(v_1),\ldots,df(v_s)).$$

For the purpose of differentiation, we will use ϕ^*_t.

Now, we define the *Lie derivative* of a tensor field A with respect to a vector field X:

$$L_X A = \lim_{\epsilon\to 0}\frac{1}{\epsilon}(\phi^*_\epsilon(A_{\phi_\epsilon(p)}) - A_p).$$

The Lie derivative has two very pertinent properties: $L_X(f) = Xf$ and $L_X(Y) = [X,Y]$ for $X, Y \in \mathfrak{X}(M)$ and $f \in \mathcal{F}(M)$. We will use these properties as a guide in generalizing tensor derivatives.

Returning to [2], let us axiomatize the properties of the Lie derivative in order to explore it in more generality. Define a *tensor derivation* as a class of maps $D: \mathfrak{X}(M) \times \mathcal{T}^r_s(M) \to \mathcal{T}^r_s(M)$ such that

(1) D is $\mathbb{R}$-linear in $\mathcal{T}^r_s(M)$,
(2) D satisfies $D(X, A\otimes B) = D(X,A)\otimes B + A\otimes D(X,B)$, and
(3) D commutes with contractions, i.e., $D(X, C(A)) = C(D(X,A))$ for any contraction C.

The first argument, X, is generally assumed. We continue the discussion with X fixed in $\mathfrak{X}(M)$.

Tensor derivations have the following property: if $A \in \mathcal{T}^r_s(M)$, then

$$\begin{aligned} D[A(\theta^1,\ldots,\theta^r,X_1,\ldots,X_s)] &= (DA)(\theta^1,\ldots,X_s) \\ &+ \sum_{i=1}^{r} A(\theta^1,\ldots,D\theta^r,\ldots,X_s) \\ &+ \sum_{i=1}^{s} A(\theta^1,\ldots,DX_j,\ldots,X_s). \end{aligned}$$

That is, A is evaluated on r one-forms and s vector fields A and then D is applied to the resulting *function*, the result is characterized by the right-hand side of the equation. This is useful in solving for $DA(\cdots)$ if $D(A(\cdots))$ and $\sum A(D\cdots)$ are known.

Note: it follows almost immediately that if D_1 and D_2 are two tensor derivations that agree exactly on functions and vector fields, then they must be the same. For if A is an arbitrary tensor, we can characterize DA with arbitrary arguments in terms of D of those arguments and D of A with those arguments (which is a function). If we have agreement on functions, vector fields, *and forms*, it follows. But $D\theta(X) = D(\theta(X)) - \theta(DX)$, and since on the right hand side we have on D of a function and D of a vector field, functions and vector fields suffice.

Conversely, if we have information about a derivation-like map on functions and vector fields, we can build a tensor. If there is a vector field $V \in \mathfrak{X}(M)$ and a mapping $\delta: \mathfrak{X}(M) \to \mathfrak{X}(M)$ such that $\delta(fX) = (Vf)X + f\delta(X)$ for all $f \in \mathcal{F}(M)$, then there is a *unique* tensor derivation D on M such that $Df = Vf$ and $DX = \delta(X)$. The proof can be discovered by noting the characterization on one-forms, extending to (r, s) tensors via the above formula, then showing that it commutes with contractions.[18]

The Lie derivative can be *defined* by the properties established above: L_X is the unique tensor derivation such that $L_X f = Xf$ and $L_X Y = [X, Y]$. Uniqueness follows from the previous proposition because the two conditions characterize L_X on $\mathcal{F}$ and $\mathfrak{X}$.

In fact, Lie derivatives are intimately related to more general tensor derivations. To wit: as related in [4], p. 30, for any tensor derivation D, there are unique vector field X and type-$(1,1)$ tensor S such that $D = L_X + S$. This is the case because, since D is a derivation when restricted to $\mathcal{F}(M)$, there is a unique vector field X whose operation coincides with D on $\mathcal{F}(M)$. Thus, $D - L_X$ is a derivation on $\mathcal{T}(M)$ which is zero on $\mathcal{F}(M)$. Put $E = D - L_X$. If Y is a vector field and f a function, then EY is a vector field and $E(fY) = (Ef)Y + fEY = fEY$ since $Ef = 0$, which is a vector field. Thus, E is an $\mathcal{F}$-linear map on $\mathfrak{X}(M)$, so it can be interpreted uniquely as a $(1,1)$ tensor. Since E and $D - L_X$ coincide on functions and vector fields, they are the same. Therefore, $D = L_X + E$, where L_X is the Lie derivative with respect to some $X \in \mathfrak{X}(M)$ and $E \in \mathcal{T}_1^1(M)$.

3. Forms, Exterior Derivatives, and the de Rham Cohomology

3.1. **Exterior differentiation.**

We follow [6] in the development of the exterior derivative. Since we can regard $\mathcal{F}(M)$ as equivalent to $\mathcal{T}_0^0(M)$, or 0-forms, we would like to see which properties of functions generalize to forms. Recall the differential of a function: $f \in \mathcal{F}(M)$ gives us, at each point $p \in M$, the one-form df. When we first defined the differential, we were considering functions $f : M \to N$. In the case of $\mathcal{F}$, however, the image is $\mathbb{R}$; the tangent space at each point of $\mathbb{R}$ is again $\mathbb{R}$, so we can think of the function df on the tangent space as mapping tangent vectors to $\mathbb{R}$ — i.e., a one-form.

We therefore may write df in terms of the basis elements of $\mathcal{T}_1^0$. The i^{th} coefficient of df is $df(\partial_i)$. But $df(\partial_i) = \frac{\partial f}{\partial x^i}$, so

$$df = \sum \frac{\partial f}{\partial x^i} dx_i.$$

Thus, $d : \mathcal{F}(M) \to \mathcal{T}_1^0(M)$. This can be generalized inductively to arbitrary forms. Define the *exterior derivative* $d : \Lambda(M) \to \Lambda(M)$ by its action on products:

$$d(\omega \wedge \phi) = (d\omega) \wedge \phi + (-1)^k \omega \wedge (d\phi),$$

where $\omega \in \Lambda^k(M)$. Note that this is consistent with 0-forms, since if $k = 0$ and $f, g \in \mathcal{F}$, then $d(fg) = df \wedge g + (-1)^0 f \wedge dg$ when interpreting the wedge product on $\Lambda^0(M)$ as function multiplication.

We can gain some understanding by considering coordinates. If ω is a k-form, we can write ω as a linear combination of k-wedges of dx_i with coefficients in $\mathcal{F}(M)$. Then d is defined in terms of this basis by

$$d\omega = \sum_i df_i \wedge dx_{i_1} \wedge dx_{i_2} \wedge \ldots \wedge dx_{i_k}.$$

This is independent of coordinate system. The image is a $k+1$-form with coefficients that are partial derivatives of the f_i. (Note that this is consistent with setting $\Lambda^m = \{0\}$ for $m > n$.)

---

[18] It is tedious but not difficult.

The map $d^2 : \Lambda^k \to \Lambda^{k+2}$ given by repeated exterior differentiation is identically zero. Without loss of generalization, set $\omega = f dx_1 \wedge dx_2 \wedge \ldots \wedge dx_k$.[19] Then

$$d\omega = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} dx_i \wedge dx_1 \wedge \ldots \wedge dx_k = \sum_{i=k+1}^{n} \frac{\partial f}{\partial x^i} dx^i \wedge dx_1 \wedge \ldots \wedge dx_k$$

since for $i = 1, \ldots, k$ the repetition of dx_i kills the term. Applying the exterior derivative again,

$$d^2\omega = \sum_{i=k+1}^{n} \sum_{j=k+1}^{n} \frac{\partial^2 f}{\partial x^i \partial x^j} dx_j \wedge dx_i \wedge dx_1 \wedge \ldots \wedge dx_k.$$

There are now two cases: $i = j$ and $i \neq j$. In the first, $dx_i \wedge dx_j = 0$. In the second, there are always *exactly two* terms in the double sum with the same coefficient (order of partials does not matter), dx_i, and dx_j: $dx_j \wedge dx_i \wedge dx_1 \wedge \ldots \wedge dx_k$ and $dx_i \wedge dx_j \wedge dx_1 \wedge \ldots \wedge dx_k$. But the two are opposites, since $dx_i \wedge dx_j = -dx_j \wedge dx_i$. So we can always pair off the nonzero terms in the sum cancel them. Hence, $d^2\omega = 0$.

Exterior differentiation is linear, so a term-by-term killing-off results in any arbitrary k-form being mapped to zero by d^2. In fact, when we regard d as an operator on the graded algebra Λ, it retains the property that $d^2 = 0$. This leads us to ask: which k-forms can be written as the exterior derivative of $(k-1)$-forms - that is, which k-forms are *exact*? And which k-forms will map to zero upon application of the exterior derivative — that is, which k-forms are *closed*? What is the difference between the closed forms and the exact forms? The answers are *topological.*

The fact that $d^2 = 0$ means that we can employ homology theory. In our case, the study of the homology groups between the k-forms is known as *de Rham cohomology.*

3.2. **Homology.** This section is adapted from a paper I wrote for the Geometric Topology class developing homology theory and applying it to the Poincare homology sphere. The section draws on [7], Sections 41-44.

3.2.1. *Definitions.* The goal of this subsection is to find groups that describe some properties of topological spaces. We motivate this by noting that any locally Euclidean space can be partitioned into triangles, or higher-dimensional analogues of triangles, so any general properties of triangulations are applicable to them.

We call the set

$$\{(x_1, \ldots, x_{n+1}) | x_1 + \ldots + x_{n+1} = 1, x_i \geq 0 \text{ for all } i = 1, \ldots, n+1\} \subseteq \mathbb{R}^{n+1}$$

the *standard geometric n-simplex.* A *geometric n-simplex* is a subset of $\mathbb{R}^{n+1}$ that is affine-equivalent to the standard geometric n-simplex. We are justified in speaking of a geometric n-simplex as a topological object by taking the subspace topology relative to the topology of $\mathbb{R}^{n+1}$.

A *combinatorial n-simplex* is a set of $n+1$ points. We can identify geometric n-simplices and combinatorial n-simplices. To get a combinatorial n-simplex from a geometric n-simplex, read off the vertices of the geometric n-simplex in the direction corresponding to the orientation. To get a geometric n-simplex from a combinatorial n-simplex, identify P_i as e_i, create

[19] We do this for the sake of notation; this *is* the general case, just with convenient relabeling.

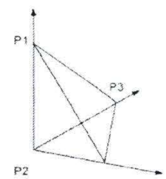

FIGURE 1. The standard geometric 2-simplex in $\mathbb{R}^3$.

a closed path by following straight lines from point to point of the combinatorial n-simplex, and identify the triangle created by filling in the closed path.

Using this correspondence, unless rigor demands otherwise, from now on we simply speak of "n-simplices" and understand that we may switch between combinatorial and geometric simplices at will.

A *sub-simplex* of an n-simplex is some lower-dimensional simplex within the n-simplex. For example, a face of an oriented tetrahedron is a sub-simplex of the tetrahedron itself. An edge of the tetrahedron is a sub-simplex of the tetrahedron. If $T = P_1P_3P_2P_4$ is a 3-simplex, $P_1P_2P_4$ is a two-dimensional sub-simplex of T.

A *simplicial complex* is a collection of n-simplices glued together along *entire* points, faces, edges, or their higher-dimensional analogues. That is, for any two n-simplices in a simplicial complex, their intersection is either empty or a k-dimensional simplex corresponding to some k-dimensional sub-simplex of *each* of the two simplices. Note that simplicial complexes need not be connected.

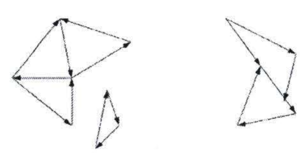

FIGURE 2. On the left above is a 2-dimensional simplicial complex. The construction on the right is not a simplicial complex.

We can place a topology on the complex by, in the interior of each n-simplex, using the subspace topology we identified above, and at the intersection of each simplex, using the quotient topology.

Note that, after affine motion, there are only two possible geometric n-simplices, which correspond to construction in right-handed and left-handed coordinate systems. We can also put an orientation on combinatorial n-simplices: let the points be traversed in some order. Then make any two n-simplices which have the same sign be equivalent. That is, given $n+1$ points, the two combinatorial n-simplices we can make with those points correspond to the two cosets of S_{n+1}/A_{n+1}.

If X is an m-dimensional simplicial complex, a *simplicial n-chain*, $0 \leq n \leq m$, is a formal sum of finitely many n-dimensional simplices of X. We denote the set of n-chains of X by $C_n(X)$; it is the free abelian group generated by the n-dimensional simplices of X.

We can relate the sets of chains of X with the *n-dimensional boundary operator* $\partial_n : C_n(X) \to C_{n-1}(X)$, which is defined on the generators of $C_n(X)$ by

$$\partial_n(P_1P_2P_3\cdots P_{n+1}) = \sum_{k=1}^{n+1}(-1)^{k+1}P_1\cdots P_{k-1}\hat{P}_kP_{k+1}\cdots P_{n+1},$$

where the circumflex denotes omission from the formal product.

As its name notes, the boundary operator maps a simplicial complex to its boundary. For example, the boundary of an oriented triangle $P_1P_2P_3$ is the chain $\partial_2(P_1P_2P_3) = P_2P_3 - P_1P_3 + P_1P_2$. This corresponds to starting at P_2, going to P_3, then following P_3 to P_1 (i.e., the negative of P_1P_3), then from P_1 back to P_2.

3.2.2. *Cycles, Boundaries, Homology Group.* Given a simplicial complex X, we have a set of abelian groups $\{C_k(X)\}$ related to X and homomorphisms ∂_k between these groups. It is only natural to ask about the images and kernels of these maps.

Let us consider $\partial_n : C_n \to C_{n-1}$. We have defined the boundary of a simplicial chain as the chain of boundaries of its terms; boundaries of generators are what we would expect them to be, respecting orientation. So a chain with zero boundary is a chain for which the chain of the boundaries of its terms cancels — in some sense, it is a chain that *closes upon itself.* The chains that close on themselves, which make up $Ker(\partial_n)$, are called the *n-cycles of* X, and we denote $Z_n = Ker(\partial_n)$.

FIGURE 3. This 2-simplex is a 2-cycle: note that each boundary is traversed *twice*, once in each direction.

The interpretation of $Im(\partial_n)$ is more straightforward: it consists of those $(n-1)$-chains which form the boundaries of n-chains. We call these the $(n-1)$*-dimensional boundaries* and denote $B_{n-1} = Im(\partial_n)$.

We can prove a crucial result about the boundary operator. The homomorphism is defined on the basis of the chains, the simplices, so we need only consider the boundary of an n-simplex, found by following its faces from point to point. In doing so, we traverse each edge exactly twice, in opposite directions. So when we calculate the boundary of the boundary, following each edge from point to point, we see that each edge is exactly canceled: the boundary of a boundary is zero.

For example, consider the simplex $P_1P_2P_3P_4$. The boundary of this simplex is $P_2P_3P_4 - P_1P_3P_4 + P_1P_2P_4 - P_1P_2P_3$. The boundary of the *boundary* is given by $(P_3P_4 - P_2P_4 + P_2P_3) - (P_3P_4 + P_1P_4 + P_1P_3) + (P_2P_4 - P_1P_4 + P_1P_3) - (P_2P_3 - P_1P_3 + P_1P_2)$. The terms obviously cancel. The proof is easily generalized, in the style of the proof in Section 3.1 above regarding the differential operator. We symbolize this result by $\partial^2 = 0$.

As a consequence, $B_n \subseteq Z_n$ for each n. That is, $Im(\partial_n) \subseteq Ker(\partial_{n-1})$. Let us consider the three groups $C_{n+1}(X)$, $C_n(X)$, and $C_{n-1}(X)$ with boundary functions ∂_{n+1} and ∂_n:

$$C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1}.$$

B_n is normal in Z_n because Z_n is an abelian group. We may thus mod out Z_n by B_n and so we define the group $H_n = Z_n/B_n$ and name it the *n-dimensional homology group of X*.

3.2.3. *Cohomology and Homology.* Consider a sequence of boundary maps $\partial_k : C_k \to C_{k-1}$. Each boundary map induces a dual map $d_k : C^*_{k-1} \to C^*_k$, where C^*_k denotes the dual space of C_k, the group of real-valued functions on C_k. The function d_k is defined by $d_k(\omega)(c) = \omega(\partial_k(c))$, where $\omega \in C^*_{k-1}$ and $c \in C_k$.

These functions, as one might expect, themselves form a chain exactly like the homology chain; the image of d_{k+1} is denoted B^k and the kernel of d_k is Z^k. The n^{th} *cohomology group* is given by $H^n = Z^n/B^n$.

It turns out that it is very easy to calculate cohomology groups from homology groups: given H_n and H_{n-1}, to calculate H^n, merely take the abelian part of H_n and the torsion part of H_{n-1}: $H^n = FREE(H_n) \times TORSION(H_{n-1})$.

3.2.4. *Homology and Topology.* Homology and corresponding cohomology groups provide topological information about the simplicial complex over which they are defined. For example, over a simplicial complex X, H_0 provides information about whether the space is path-connected. For $Z_0 = C_0(X)$, all the points of X, and two points are in the same coset exactly when they are the boundary of a 1-chain — i.e., a path runs between them. Therefore, there is one coset for each path component of X.

3.3. **de Rham cohomology.** On a manifold M, we will use the singular homology. We will consider not actual simplices in M but rather C^∞ maps to M of simplices in $\mathbb{R}^k$. So a simplex in M is a map $\sigma : \Delta^k \to M$, where Δ^k is the k-simplex in $\mathbb{R}^{k+1}$. A simplicial complex is a set of all *real* linear combinations of (reasonably non-overlapping, touching-only-on-faces) simplices in M. So we denote by ${}_\infty S_k(M, \mathbb{R})$, or just S_k, the real vector space generated by k-simplices in M. That is, S_k is the (finite) k-chains with *real* coefficients instead of *integer* coefficients. As above, we have a boundary map $\partial_k : {}_\infty S_k \to {}_\infty S_{k-1}$ for each integer k, defined on the basis, except this time it $\mathbb{R}$-linear, not simply $\mathbb{Z}$-linear.

The differential singular homology group of M with real coefficients is the vector space

$${}_\infty H_l(M, \mathbb{R}) = Ker(\partial_k)/Im(\partial_{k+1}).$$

For more detail, see [3], pp. 141-3, and [6], p. 103-4.

We must consider integrals before proceeding. After integration is introduced, we will state the de Rham theorem.

4. Manifold Integration

We will consider the integration of forms. This is quite useful in physics — for example, in the construction of a gauge theory, one requires invariance of an integral of a form over all space under local action by a Lie group — as well as mathematically — for example, total curvature is the integral of a curvature form over a manifold. Our development of integration loosely and briefly follows [3], pp. 143-155. Additional information on integration of forms on manifolds is containted in [8], pp.174-180, and [9], Ch. 10.

4.1. **Orientation.** If V is a real n-dimensional vector space, then $\Lambda_n(V)$ has dimension 1, so $\Lambda_n(V) - \{0\}$ has two components. An *orientation* on V is a choice of components. If M is a connected differentiable manifold, M is *orientable* if there is a consistent choice of orientation on the tangent spaces. An orientation is such a consistent choice.

If M is oriented, pick $p \in M$ and let $v_1, \dots, v_n$ be a basis of $T_p(M)$ with dual basis $\delta_1, \dots, \delta_n$. The ordered basis $(v_1, \dots, v_n)$ is oriented if $\delta_1 \wedge \dots \delta_n$ is an element of the orientation (restricted to $T_p(M)$). We can in fact globally characterize orientation. The following are equivalent:

- M is orientable;
- There is a coordinate cover of M such that all pairwise Jacobian matrices have positive determinant;
- There is a nowhere-vanishing n-form on M.

4.2. **Integration.** Recall that in $\mathbb{R}^n$, if D is open, if there is a diffeomorphism $\phi : D \to \phi(D)$, if f is a bounded continuous function on $\phi(D)$, and if $A \subseteq D$, then

$$\int_{\phi(A)} f = \int_A f \circ \phi |J_\phi|,$$

where J_ϕ is the Jacobian matrix of ϕ. In other words, when you integrate a nice function after applying a diffeomorphism to the domain of integration, you get the same thing as the integral of the pullback of the function under the diffeomorphism scaled by the Jacobian of the transformation.

Since the n-forms on M are vector-space isomorphic to $\mathcal{F}(M)$ (each $\omega \in \Lambda_n(M)$ can be written as $f dx_1 \wedge \dots \wedge dx_n$), we begin integration of forms over M with integration of n-forms in particular. First, recall that under a diffeomorphism ϕ, we can define the pullback of a form ω under ϕ by $\phi^*(\omega)(v_1, \dots, v_k) = \omega(dv_1, \dots, dv_k)$. Note that ϕ^* commutes with d and preserves the wedge product. If $\omega = f\ dx_1 \wedge \dots \wedge dx_n$ is an n-form defined in a region D of $\mathbb{R}^n$, we define

$$\int_D \omega = \int_D f.$$

If $D = \phi(A)$ for some diffeomorphism ϕ, then as above ϕ induces a transformation ϕ^* on Λ_n so that

$$\int_{\phi(A)} \omega = \pm \int_A \phi^*(\omega).$$

If ϕ preserves orientation, we take $+$; if ϕ reverses orientation, we take $-$.

We are now in a position to define the integral over a singular simplex in M. If σ is a singular k-simplex in M and ω is a k-form, define

$$\int_\sigma \omega = \int_{\Delta^k} \sigma^*(\omega).$$

The extension to general singular k-chains is done by defining the integral to be linear:

$$\int_c \omega = \sum a_i \int_{\sigma_i} \omega,$$

where the a_i are coefficients of the σ_i in c.

If c is a singular $k+1$-chain in M and ω is a smooth k-form defined in a neighborhood of the image of c, then this theorem relates the propertis of c with the properties of ω:

$$\int_{\partial c} \omega = \int_c d\omega.$$

When we generalize to domains, for any subset $D \subseteq M$,

$$\int_{\partial D} \omega = \int_D d\omega.$$

This is Stokes' theorem.

4.3. **de Rham's theorem.** Stokes' theorem permits us to relate the singular homology of M with the de Rham cohomology. Let us define a linear mapping $H^k \to H_k$ from the kth de Rham cohomology group (vector space) to the dual space of the kth real differentiable singular homology group (vector space) by

$$[\alpha]([z]) = \int_z \alpha$$

for an equivalence class $[\alpha] \in H^k$ and arbitrary equivalence class $[z] \in H_k$. It is not hard to check if this is well-defined: if $\alpha, \beta \in [\alpha]$, then $\int_z (\alpha - \beta) = \int_z d\eta$ for some form η. But $\int_z d\eta = \int_{\partial z} \eta = 0$ since $\partial z = 0$. And if $w, z \in [z]$, then $\int_w \alpha - \int_z \alpha = \int_{w-z} \alpha = \int_{\partial y} \alpha = \int_y d\alpha = 0$ since $d\alpha = 0$, for some y.

The theorem says: this mapping is an isomorphism of vector spaces. The implication is that all of the topological information available via singular homology is available via differential forms and vice-versa: the study of forms on a manifold is inextricably intertwined with the study of the topology of the manifold.

4.4. **Poincare's lemma.** Here is an ingredient in the proof of de Rham's theorem: Poincare's lemma. If U is the open unit ball in $\mathbb{R}^n$, then for each $k \geq 1$, there is a linear transformation $h_k : \Lambda^k(U) \to \Lambda^{k-1}(U)$ with $h_{k+1} \circ d + d \circ h_k =$ the identity. More detail can be found in [1], p. 20; [6], p. 118; or [3], p. 155.

4.5. **The Hodge star $*$.** Both [1], pp. 35-7, and [6], pp. 150-3. If M is a manifold with a metric (see Section 5 for definition and development), the Hodge star is a linear operator $* : \Lambda^k \to \Lambda^{n-k}$ with $** = (-1)^{k(n-k)}$. On an orthonormal frame field $\{x_i\}$ (when a metric is present, one can always be gotten from a local coordinate patch by the Gram-Schmidt process), $*$ is defined as the linear operator which acts on the basis as $*(dx^1 \wedge dx^2 \wedge \ldots \wedge dx^k) = dx^{k+1} \wedge \ldots \wedge dx^n$. More generally, we define $*$ as mapping $*(dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_k}) = dx^{i_{k+1}} \wedge dx^{i_{k+2}} \wedge \ldots \wedge dx^{i_n}$, where $\{i_1, \ldots, i_n\}$ is an *even* permutation of $\{1, \ldots, n\}$. This is why we require an orientation: in order to be able to consistently choose an even permutation across all of M.

Given a vector field $V = \sum V_i \partial_i$, we can use a metric to canonically identify V with a one-form $\mathcal{V}$ by defining, in terms of action on $\mathfrak{X}$, $\mathcal{V}(W) = g(V, W)$, and denote $g(V, \cdot) \in \mathfrak{X}^*(M)$. In coordinates, $\mathcal{V} = \sum_i (\sum_j g^{ij} V_j) dx_i$. For more elaboration, see Section 5.1.

The $*$ operator is very useful in formalizing familiar vector operations — for instance, we will make heavy use of them in Section 9.1. For example, on $\mathbb{R}^3$ with the canonical coordinates, $* : \mathcal{F} \leftrightarrow \Lambda^3, \Lambda^1 \leftrightarrow \Lambda^2$ and it acts so that $*1 = dx \wedge dy \wedge dz$, $*(dx \wedge dy \wedge dz) = 1$,

$*(dx) = dy \wedge dz, *dy = -dx \wedge dz, *dz = dy \wedge dz$, and $*(dx \wedge dy) = dz, *(dx \wedge dz) = -dy$, and $*(dy \wedge dz) = dx$.

On $\mathbb{R}^3$, if $f \in \mathcal{F}$, we can represent the gradient ∇f by

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz.$$

If V is a vector field, we can represent it as a one-form via the metric; on $\mathbb{R}^3$, the metric is just $g_{ij} = \delta_{ij}$, so $V \to \mathcal{V} = V_x dx + V_y dy + V_z dz$. We have

$$\begin{aligned} *d\mathcal{V} &= * \left[\left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) dx \wedge dy + \left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \right) dx \wedge dz + \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right) dy \wedge dz \right] \\ &= \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right) dx + \left(\frac{\partial V_z}{\partial x} - \frac{\partial V_x}{\partial z} \right) dy + \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) dz. \end{aligned}$$

That is, $*d\mathcal{V}$ is a form representation of $\nabla \times V$.

Finally, since $*$ is an isomorphism, we can send $\mathcal{V}$ to $*\mathcal{V}$, a 2-form, before we apply the exterior derivative. Then we have

$$d*(V_x dx + V_y dy + V_z dz) = d(V_z dx \wedge dy - V_y dx \wedge dz + V_x dy \wedge dz) = \left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z} \right) dx \wedge dy \wedge dz.$$

Applying $*$ once more, we have the *function*

$$*d*\mathcal{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}.$$

In other words, $*d*\mathcal{V} = \nabla \cdot V$.

We also have an operator $\delta : \Lambda^k \to \Lambda^{k-1}$ given by $\delta = (-1)^{n(k+1)+1} * d*$. On functions, δ sends everything to zero. The Laplace-Beltrami operator (or Laplacian) acts on Λ^k and is defined by $\Delta = \delta d + d\delta$. (It is obviously linear.) When it acts on a function f on $\mathbb{R}^n$, $\Delta f = -\sum \partial^2 f / \partial x_i^2$. We also have an inner product induced by the $*$ operator. When the $*$ is defined in terms of the inner product, it agrees with the extension of the inner product to k-forms.

It should be emphasized that $*$ and related operators are *not well defined* without both a metric and a choice of orientation. The metric is necessary for an orthonormal basis to be chosen; the orientation is necessary so that, over a closed loop, the operator is smoothly defined (i.e., one cannot return to a starting point and have reversed sign).

5. Metrics and Curvature

This section very closely follows [2], Chapter 3, which gives a clear and reasonably complete treatment of the subject. Other sources also contain brief introductions to metrics; for example, [4], [5]. [8] gives a lower-level and less concise treatment of the subject.

Euclidean geometry concerns itself with *lengths* and *angles*. In[20] $\mathbb{R}^n$, we can do any geometry with the dot product, even analytic geometry. Since there is an isomorphism between the whole of $\mathbb{R}^n$ and the tangent space at any particular point ($v = (v_1, \ldots, v_n) \mapsto \sum v_i \partial_i$), on a manifold we can introduce a dot product on tangent spaces. That lets us do geometry, roughly speaking, infinitesimally — for example, the work done by a (vector) force field X on a smooth curve $\gamma : [0, 1] \to \mathbb{R}^n$ is a geometric notion, force dot distance, found

[20]This is aptly termed *Euclidean* space.

by adding up the "infinitesimal distance" dr traveled at each point dotted into the force at each point. Then, since $dr = \gamma' dt$, and if $\theta(t)$ is the angle between force and velocity at time t,

$$W = \int F \cdot dr = \int_0^1 F(\gamma(t)) \cdot (\gamma'(t)) dt = \int_0^1 \|F(\gamma(t))\| \|\gamma'(t)\| \cos(\theta(t)) dt.$$

Let us generalize the properties of the Euclidean metric. The dot product, or inner product, is a symmetric tensor of type $(0, 2)$, a *bilinear form* or *metric tensor* or, simply, *metric*[21]. A tensor field is *symmetric* when the exchange of any two inputs does not change the output. We here restrict our attention to nondegenerate symmetric bilinear forms, the discipline of pseudo-Riemannian geometry; antisymmetric bilinear forms are the domain of symplectic geometry.[22]

If M is a manifold equipped with a metric tensor g, under an arbitrary coordinate system x_i we can write $g = \sum g_{ij} dx^i \otimes dx^j$, where $g_{ij} = g(\partial_i, \partial_j)$. When applied to two vector fields, the metric acts as a function $g : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathcal{F}(M)$ defined by $g(V, W) = \sum g_{ij} V_i W_j$. Denoting the matrix $G = (g_{ij})$, this is equivalent to matrix multiplication: $g(V, W) = V^T G W$. For example, in two dimensions,

$$g(V, W) = \begin{pmatrix} V_1 & V_2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix}.$$

Because g is nondegenerate, in every coordinate system the associated matrix G is invertible; the components of G^{-1} are denoted by g^{ij}.

Following the strategy of *defining* the properties of the inner product in order to explore the general consequences, we define the length or norm of a tangent vector v by $\|v\|^2 = g(v, v)$. Two tangent vectors are orthogonal if $g(v, w) = 0$. A collection $\{v_i\}$ of tangent vectors is orthonormal if $g(v_i, v_j) = \delta_{ij}$. With this generalization, we are now able to construct and express more general geometries.

We can identify surfaces as geometrically equivalent. If we have two pseudo-Riemannian manifolds M and N with metrics g_M and g_N, we say a map $\phi : M \to N$ is an *isometry* if ϕ is a diffeomorphism preserving the metric tensor — that is, if $g_N(d\phi(v), d\phi(w)) = g_M(v, w)$ for each v, w in $T(M)$. Since ϕ is a diffeomorphism, it preserves differential structure; being an isometry is much, much stricter, since distance and angle are preserved as well.

5.1. **Covariant differentiation.** We now wish to explore some of the intrinsic geometry of a manifold equipped with a metric. We start by identifying a particular tensor derivative. On a Euclidean or semi-Euclidean space, if $V, W \in \mathfrak{X}(\mathbb{R}^n)$, the *covariant derivative of* V *with respect to* W is

$$D_V W = \sum V(W^i) \frac{\partial}{\partial x_i}.$$

Here are two facts about the natural covariant derivative on Euclidean space. First, $[X, Y] = D_X Y - D_Y X$; second, $X\langle Y, Z\rangle = \langle D_X Y, Z\rangle + \langle Y, D_X Z\rangle$. To see the first, recall that

$$[X, Y] = \sum_i \sum_j \left(V^j \frac{\partial W^i}{\partial x_j} - W^j \frac{\partial V^i}{\partial x_j} \right) \frac{\partial}{\partial x_i}.$$

[21] Not to be confused with the distance map from analysis of the same name.

[22] Don't get "bilinear forms" mixed up with "2-forms;" they often have entirely different properties and implications!

The relationship is merely a matter of applying the definition. The second is a matter of tedious calculation.

It is not particularly helpful that this definition uses $\mathbb{R}^n$'s canonical coordinates, so, following our theme of axiomatizing and generalizing, we define a *connection* D on a manifold M as a function $D : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ such that

- $D(fV + W, X) = fD(V, X) + D(W, X)$,
- $D(V, aW) = aD(V, W)$, and
- $D(V, fW) = (Vf)W + fD(V, W)$ for $f \in \mathcal{F}(M)$.

To signify the fact that D is $\mathcal{F}$-linear in the first argument but only $\mathbb{R}$-linear in the second, we will denote the connection evaluated on two vector fields by $D(V, W) = D_V W$, and call it the covariant derivative of W with respect to V for the connection D.

We want to build a connection D on a general semi-Riemannian manifold that generalizes the properties of the Euclidean covariant derivative — namely, such that for all $V, W, X \in \mathfrak{X}(M)$, $[V, W] = D(V, W) - D(W, V)$ and $Vg(W, X) = g(D_V W, X) + g(W, D_V X)$. To do this, we first note a generally useful fact: the introduction of a metric g on a manifold M induces a canonical $\mathcal{F}(M)$-linear isomorphism between $\mathfrak{X}(M)$ and $\mathfrak{X}^*(M)$. If $V \in \mathfrak{X}(M)$, identify it with the one-form V^* which acts on a vector field X by $V^*(X) = g(V, X)$. Of course, the metric, being a tensor, is $\mathcal{F}$-linear in both arguments, so V^* is a one-form and the identification is also $\mathcal{F}$-linear. So it is a vector space homomorphism. To see it is an isomorphism, we need only show bijectivity. If $g(V, X) = g(W, X)$ for each X, then $g(V - W, X) = g(V, X) - g(W, X) = 0$. But since g is nondegenerate, the only vector field which maps to zero with *every* vector field is the zero vector field. Thus, the identification is injective. And every one-form θ corresponds such a vector field V; in coordinates, write $\theta = \sum \theta_i dx^i$ and put $V = \sum_i \sum_j g^{ij}\theta_i \partial_j$ (this is the operation of *lowering indices*). To see if θ agrees with $g(V, \cdot)$, need only check indices; applying $\mathcal{F}$-linearity, $g(V, \partial_k) = \sum_i \sum_j g^{ij}\theta_i g(\partial_j, \partial_k) = \sum_i \sum_j g^{ij} g_{jk} \theta_i$. The two matrices are inverses, so $= \sum_i \theta_i \delta_{ik} = \theta_k = \theta(\partial_k)$.

We can transform a vector field into a one form, and vice-versa. Corresponding pairs are *metrically equivalent*; physicists regard vectors as contravariant tensors of rank one and one-forms as covariant tensors of rank one, and do not distinguish between them.

With this information, we can now show not just the existence but the uniqueness of a connection D with the two further properties identified above:

- $[V, W] = D_V W - D_W V$
- $Xg(V, W) = g(D_X V, W) + g(V, D_X W)$.

In fact, D is entirely characterized by the *Koszul formula*

$$
\begin{aligned}
2g(D_V W, X) &= Vg(W, X) + Wg(X, V) - Xg(V, W) \\
&\quad - g(V, [W, X]) + g(W, [X, V]) + g(X, [V, W]).
\end{aligned}
$$

The proof is not difficult;[23] we sketch it, following [2], pp. 59-61. First, uniqueness: we show D, if it exists, is consistent with the Koszul formula. If D satisfies the two axioms, then, term by term, $Vg(W, X) + Wg(X, V) - Xg(V, W) - g(V, [W, X]) + g(W, [X, V]) + g(X, [V, W]) = [g(D_V W, X) + g(W, D_V X)] + [g(D_W X, V) + g(X, D_W V)] - [g(D_X V, W) +$

[23]One expects the *deep* part was dreaming up the Koszul formula.

$g(V, D_X W)] - [g(V, D_W X) - g(V, D_X W)] + [g(W, D_X V) - g(W, D_V X)] + [g(X, D_V W) - g(X, D_W V)] = 2g(D_V W, X)$.

Now, for existence, set $F(V, W, X)$ to be the right hand side of the Koszul formula, show it is $\mathcal{F}$-linear in X for fixed V, W (and so is a one-form). By the pairing between vectors fields and one-forms, there is a vector field corresponding to $2F(V, W, \cdot)$; call it $D_V W$. From this vector field, deduce the three conditions of a connection and the two additional conditions.

The connection D is called the *Levi-Civita connection.*

Let a coordinate system on M be fixed; the *Christoffel symbols* of the coordinate system are the functions Γ_{ij}^k such that

$$D_{\partial_i} \partial_j = \sum_k \Gamma_{ij}^k \partial_k.$$

That is, the Christoffel symbols are the components in the coordinate system of covariant differentiation of the natural coordinate system vector fields. It is trivial to verify that $[\partial_i, \partial_j] = 0$, so $\Gamma_{ij}^k = \Gamma_{ji}^k$.

Note that because D itself is not a tensor, the Christoffel symbols do not transform like tensors (i.e., they are not coordinate expressions of a $(1, 2)$ tensor).

Using the Christoffel symbols, let's build up a coordinate description of how the Levi-Civita covariant derivative operates on vector fields. Let $W = \sum W^j \partial_j$; then

$$\begin{aligned} D_{\partial_i} W = \sum D_{\partial_i}(W^j \partial_j) &= \sum (\partial_i (W^j) \partial_j + W^j D_{\partial_i} \partial_j) \\ &= \sum_j \frac{\partial W^j}{\partial x_i} \partial_j + \sum_j \sum_k \Gamma_{ij}^k W^j \partial_k \\ &= \sum_k \left[\frac{\partial W^k}{\partial x_i} + \sum_j \Gamma_{ij}^k W^j \right] \frac{\partial}{\partial x_k} \end{aligned}$$

after cleverly reindexing.

The Christoffel symbols themselves can be expressed solely in terms of the metric. In the Koszul formula, if $V = \partial_i$, $W = \partial_j$, and $X = \partial_m$, after a flurry of algebra,

$$\Gamma_{ij}^k = \frac{1}{2} \sum_m g^{km} \left[\frac{\partial g_{jm}}{\partial x^i} + \frac{\partial g_{im}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^m} \right]$$

The Levi-Civita covariant derivative can be extended to arbitrary tensors: in fact, by the theorem in Section ?? on constructing a general tensor derivations based only on its action on vector fields and functions, there is a unique tensor derivation on M such that $D_V f = Vf$ for smooth functions and $D_V W$ is the Levi-Civita covariant derivative for all vector fields W.

A tensor field A is *parallel* if, for all vector fields X, $D_X A = 0$. In some sense, Christoffel symbols measure the deviation of coordinate vector fields from parallelity. For example, a zero vector field is always parallel; in $\mathbb{R}^3$, any vector field with constant coefficients is a parallel vector field.

For functions, $df = \sum \frac{\partial f}{\partial x_i} dx^i$ conveniently collects all of the partial derivatives of f. We can generalize the notion of the differential to arbitrary tensors: for $A \in \mathcal{T}_s^r(M)$, define

$$(DA)(\theta^1, \ldots, X_s, V) = (D_V A)(\theta^1, \ldots, X_s).$$

5.2. **Geodesics.** If we have a curve $\gamma : I \to M$ with a vector field Z over it (i.e., a vector field $Z \in \mathfrak{X}(M)$ restricted to $\gamma(I)$; vector fields restricted to $\gamma(I)$ are denoted $\mathfrak{X}(\gamma)$), we can define the vector field's rate of change with respect to the curve's parameter. This is the *induced covariant derivative* $Z' = DZ/dt : \mathfrak{X}(\gamma) \to \mathfrak{X}(\gamma)$. With respect to arbitrary coordinates,

$$Z' = \sum \frac{dZ^i}{dt}\partial_i + \sum Z^i D_{\gamma'}\partial_i,$$

in other words, the sum of how quickly the coordinate functions of Z change and how quickly the direction of Z changes. Writing the covariant derivative in terms of Christoffel symbols,

$$Z' = \sum_k \left[\frac{dZ^k}{dt} + \sum_i \sum_j \Gamma^k_{ij} \frac{d(x_i \circ \gamma)}{dt} Z^j\right] \partial_k.$$

If $Z' = 0$, then Z is *parallel* on α.

There is a special subset of curves γ such that the velocity field γ' is parallel on γ; we term such a curve a *geodesic*. If $\{x_i\}$ is a coordinate system on M, then we can express $\gamma' = \sum_l \frac{d(x_l \circ \gamma)}{dt}\partial_l$. By the above, setting $Z = \gamma'$, we must have that

$$0 = Z' = \sum_k \left[\frac{d}{dt}\frac{d(x_k \circ \gamma)}{dt} + \sum_i \sum_j \Gamma^k_{ij} \frac{d(x_i \circ \gamma)}{dt}\frac{d(x_j \circ \gamma)}{dt}\right] \partial_k,$$

which happens exactly when the coordinate functions are zero, i.e., γ satisfies the system of differential equations

$$\frac{d^2(x_k \circ \gamma)}{dt^2} + \sum_i \sum_j \Gamma^k_{ij} \frac{d(x_i \circ \gamma)}{dt}\frac{d(x_j \circ \gamma)}{dt} = 0$$

for each k.

By the existence and uniqueness theorem for solutions of ordinary differential equations, we have the local existence of geodesics: if $v \in T_p(M)$, there is an interval $I \ni 0$ and a unique geodesic $\gamma : I \to M$ with $\gamma'(0) = v$. Here, we have $\gamma(0) = p$, so we call γ a geodesic starting at p with initial velocity v.

Much like integral curves, if two geodesics agree in position and speed at a point, they are the same. This uniqueness means that for every tangent vector at a point, there is a unique *maximal* geodesic which goes through that point with that velocity, much like the uniqueness of maximal integral curves.

The parametrization of a curve is significant since being geodesic depends on the velocity of the curve. In fact, a reparametrization $\gamma \circ h$ of γ is a geodesic if and only if h is linear in t. Physical intuition may be helpful. If one thinks of a geodesic as representing unconstrained motion, traveling along a geodesic at a different speed is impossible, for then one would have to be constantly correcting one's motion — i.e., accelerating.

5.3. **Curvature.** For a surface embedded in $\mathbb{R}^n$, we can measure the surface's deviation from planarity by checking the rate of change of a normal vector with respect to the coordinate directions. This deviation is a measure of curvature; a flat space has zero curvature, a saddle-shaped space has negative curvature, and a ball-shaped space has positive curvature.

By Gauss' *theorema egregium* ("remarkable theorem")[24], extrinsically defined curvature is actually an entirely intrinsic property. So, as always, wishing to study the abstract properties of curvature, we axiomatize those properties and explore their consequences. A further motivation for this is the fact that Lie derivatives satisfy $L_{[X,Y]} = [L_X, L_Y]$ $(= L_X L_Y - L_Y L_X)$. If $[X, Y] = 0$, then the operators L_X and L_Y commute with each other. However, in general the covariant derivative does *not* satisfy $D_{[X,Y]} = [D_X, D_Y]$. We can measure the failure; and the extent to which $D_{[X,Y]}$ deviates from $[D_x, D_Y]$ is intimately related to the curvature.

Therefore, if M is a pseudo-Riemannian manifold with metric g and Levi-Civita connection D, the function $R : \mathfrak{X}(M)^3 \to \mathfrak{X}(M)$ defined by

$$R(X, Y, Z) = R_{XY}Z = D_{[X,Y]}Z - [D_X, D_Y]Z$$

is a $(1, 3)$ tensor field on M called the *Riemannian curvature tensor* on M.

If R is $\mathcal{F}$-multilinear, this interpretation holds because, for any one-form θ, we can write $R(\theta, X, Y, Z) = \theta(R(X, Y, Z))$. And R is indeed $\mathcal{F}$-multilinear; $\mathbb{R}$-mulilinearity is obvious, so we need only show that $R(fX, gY, hZ) = fghR(X, Y, Z)$, which follows from $[fX, Y] = XfY + f[X, Y]$ and the $\mathcal{F}$-linearity in W of $D_V W$.

Pointwise, if we fix two tangent vectors x, y at p, $R_{xy} : T_p(M) \to T_p(M)$ by $z \mapsto R_{xy}z$. This is the *curvature operator*. Here are some properties:

- $R_{xy} = -R_{yz}$
- $g(R_{xy}v, w) = -g(R_{xy}w, v)$
- $R_{xy}z + R_{yz}x + R_{zx}y = 0$
- $g(R_{xy}v, w) = g(R_{vw}x, y)$

for each $w, x, y, z \in T_p(M)$. For proof, see [2], p. 75. The first two identities identify skew-symmetry in the curvature operator. The third identity, the *first Bianchi identity*, bears family resemblence to the Jacobi identity, $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ for vector fields X, Y, Z, which can be used to prove the second Bianchi identit, concerning the covariant differential of the Riemann tensor: at a point p with tangent vectors x, y, z, $(D_z R)(x, y) + (D_x R)(y, z) + (D_y R)(z, x) = 0$.

Now, having stated some abstract properties of the Riemann curvature tensor, let us write it in coordinates. We establish the components[25] by $R_{\partial_k \partial_l}\partial_j = \sum_i R^i_{jkl}\partial_i$. By the definition of the tensor, $R_{\partial_k \partial_l}\partial_j = D_{\partial_l}(D_{\partial_k}\partial_j) - D_{\partial_k}(D_{\partial_l}\partial_j)$ since the bracket of two coordinate vector fields is identically zero. Cranking through the coordinates of the covariant derivative and equating coordinate functions, we have that

$$R^i_{jkl} = \left(\frac{\partial}{\partial x_l}\Gamma^i_{kj} - \frac{\partial}{\partial x_k}\Gamma^i_{lj}\right) + \sum_m \left(\Gamma^i_{lm}\Gamma^m_{kj} - \Gamma^i_{km}\Gamma^m_{lj}\right).$$

This formula is difficult to use. O'Neill recommends using other characteristics of M to guide calculations.

We can simplify our approach to use two-dimensional subspaces to the tangent space, or tangent planes. If $v, w \in T_p(M)$, define $Q(v, w) = g(v, v)g(w, w) - g(v, w)^2$. A two-dimensional subspace Π of $T_p(M)$ is nondegenerate if and only if $Q(v, w) \neq 0$ for each basis

[24] My wife, who has a degree in Latin (*inter alia*), checked this; *theorema* is a neuter 3^{rd} declension noun, and so *egregium* is neuter to match it.

[25] Be careful about the order of the indices!

v, w of Π. If Π is a nondegenerate tangent plane,

$$K(\Pi) = g(R_{vw}v, w)/Q(v, w),$$

called the *sectional curvature of* Π, is independent of the choice of basis v, w of Π. This is true because if a second basis x, y is given of Π, there is a transformation matrix A between $\{x, y\}$ and $\{v, w\}$. By direct computation, $g(R_{vw}v, w) = \det(A)^2 g(R_{xy}x, y)$ and $Q(v, w) = \det(A)^2 Q(x, y)$.

In fact, K in some sense determines R. If $K(\Pi) = 0$ at p for each plane in $T_p(M)$, then $R = 0$ at p. If $F : T_pM^4 \to \mathbb{R}$ is a function which has the four symmetries outlined above, define

$$K(v, w) = \frac{F(v, w, v, w)}{g(v, v)g(w, w) - g(v, w)^2},$$

for all non-colinear $v, w \in T_p(M)$, then $g(R_{vw}x, y) = F(v, w, x, y)$. To see this, use the previous sentence by putting $\Delta = g(R_{vw}x, y) - F(v, w, x, y)$ and checking that the corresponding K is zero for all linearly independent pairs v, w.

5.4. **Frame fields.** Orthonormal bases, or *frames*, for tangent planes have several nice properties.[26] These can be extended to orthonormal vector fields, or *frame fields* (also called *tetrads* or *vierbeins*.) There is often not a *global* orthonormal frame field, but they always exist locally (given a frame on a tangent plane, extend it via geodesics to a neighborhood). If a frame field exists, then any vector field can be expressed as a linear combination of the frame fields by orthonormal expansion ($V = \sum g(V, E_i)E_i$). Then, in terms of the frame field,

$$\begin{aligned} g(V, W) &= g(\sum g(V, E_i)E_i, \sum g(W, E_j)E_j) \\ &= \sum_i \sum_j g(V, E_i)g(W, E_j)g(E_i, E_j) = \sum g(V, E_i)g(W, E_i) \end{aligned}$$

by orthonormality.

It turns out that, if $\gamma : [0, 1] \to M$ is a curve in M and $\{e_i\}$ is an orthonormal set of vectors at $\gamma(0)$, there is a unique parallel frame field $E_i \in \mathfrak{X}(\gamma)$ with $E_i(0) = e_i$. Neither this proposition nor the previous one is obvious; for proof, see [2], pp. 84-5.

5.5. **Ricci and scalar curvature.** If R is the curvature tensor of M, then the *Ricci tensor* of M is the unique[27] (up to sign) contraction of R. Abstractly, $Ric = C_3^1(R) \in \mathcal{T}_2^0(M)$. Relative to a coordinate system, $R_{ij} = \sum R^m_{ijm}$. Relative to a frame field E_i, the Ricci tensor is given by $Ric(X, Y) = \sum_m g(R_{XE_m}Y, E_m)g(E_m, E_m)$. The Ricci tensor is entirely determined by sectional curvature: if e_i is a frame at p with $u = e_1$, then $Ric(u, u) = \sum_i g(e_i, e_i)g(R_{ue_i}(u), e_i) = g(u, u)\sum K(u, e_i)$, or, it is the sum of any $n - 1$ orthogonal nondegenerate planes through S.

If we contract the Ricci tensor, then, we get scalar curvature $S = C(Ric)$. This is, of course, a C^∞ function. Coordinatewise, $S = \sum g^{ij}R_{ij}$. Relative to a frame field E_i, $S = \sum' K(E_i, E_j) = 2\sum_{i<j} K(E_i, E_j)$. Also, we have $dS = 2div(Ric)$, where div is the divergence operator on tensors. For more detail, see [2], pp. 87-9.

---

[26]Among them, orthonormality

[27]Compare to the symmetries of R – Ric is the only nonzero one.

6. Lie Theory

This section is a brief summary of the basic tenets of Lie groups; much more information can be found in [3], Ch. 3, and [10].

In physical applications, we are going to be interested in constructing quantities that are invariant under groups of local transformations. Such groups are examples of *Lie groups.*

A *Lie group* G is a differentiable manifold that is also a group such that the map $G\times G \to G$ given by $(\sigma, \tau) \mapsto \sigma\tau^{-1}$ is C^∞. A Lie algebra $\mathfrak{g}$ over $\mathbb{R}$ is a real vector space with a bilinear bracket operation such that, for each $x, y, z \in \mathfrak{g}$, $[x, y] = -[y, x]$ and $[[x, y], z] + [[y, z], x] + [[z, x], y] = 0.$[28]

The smooth group operation on G adds an additional layer of structure to the manifold that we can profitably exploit. We define *left* and *right translation by* σ to be the diffeomorphisms $l_\sigma(\tau) = \sigma\tau$ and $r_\sigma(\tau) = \tau\sigma$, respectively. A vector field X on G, which we do not initially assume to be smooth, is *left invariant* if, for each $\sigma \in G$, X is left-related to itself – i.e.,

$$dl_\sigma \circ X = X \circ l_\sigma.$$

If G is a Lie group, set $\mathfrak{g}$ to be its set of left-invariant vector fields. Then

- $\mathfrak{g}$ is a real vector space and $\alpha : \mathfrak{g} \to T_eG$ given by $\alpha(X) = X(e)$ is a vector space isomorphism;
- Left invariant vector fields are smooth;
- The Lie bracket of two left-invariant vector fields is left invariant; and
- $\mathfrak{g}$ is a Lie algebra under the Lie bracket.

There are several equivalent interpretations of $\mathfrak{g}$. It is the Lie algebra of left-invariant vector fields on G under $[\cdot,\cdot]$, and it is also the tangent space at the identity element e of G.

Just as there are left-invariant vector fields, there are left-invariant forms: $\omega \in \Lambda^k G$ is left-invariant if $\delta l_\sigma \omega = \omega \circ l_\sigma$ for all $\sigma \in G$. (Just as with vector fields, all left-invariant forms are smooth.) We denote the left-invariant k-forms by $\Lambda^k_l(G)$, and left-invariant one-forms are known as *Maurer-Cartan forms.* Λ^1_l is the dual space of $\mathfrak{g}$.

A function $\phi : G \to H$ is a Lie group homomorphism if it is C^∞ and preserves operations. If $H = Aut(V)$ for some vector space V, i.e., $H \subseteq Gl(n, \mathbb{C})$, then we call ϕ a *representation of* G. If $\mathfrak{g}$ and $\mathfrak{h}$ are Lie algebras, then $\psi : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism if ψ is a vector space homomorphism that preserves brackets. If $\mathfrak{h} = End(V)$ for some v.s. V or $\mathfrak{gl}(n, \mathbb{R}/\mathbb{C})$, then ψ is a representation of $\mathfrak{g}$. In fact, the differential of a Lie *group* homomorphism ϕ is a Lie *algebra* homomorphism.

If $\phi : G \to H$ is a homomorphism, we can pull left invariant forms back from H to G via $\delta\phi$. The map $\delta\phi$ on one-forms is actually the *transpose of* $d\phi$: $(\delta\phi(\omega))(X) = \omega(d\phi(X))$ if $X \in \mathfrak{g}$.

The group structure locks down homomorphisms this way: if two homomorphisms ϕ, ψ have $d\phi = d\psi$ on $T_eG = \mathfrak{g}$, and G is connected, then $\phi = \psi$.

7. Vector Bundles and Sections

In this section, we draw on the development in [1], Ch. 9, and [6], Ch. 5-6. [4] also has extensive material on the subject: Ch. 2-5 of Vol. I.

[28]It is no surprise that the bracket operation on $\mathfrak{X}(M)$ with the same properties is called the "Lie bracket".

7.1. **Motivation.** To think of tangent, cotangent, tensor, and form spaces in the aggregate, we can define the *tangent bundle* $T(M) = \cup_{p\in M} T_pM$. The *projection map* $\pi : T(M) \to M$ is defined by $\pi(v) = p$ if $v \in T_pM$. Similarly, the *cotangent bundle* is $T^*(M) = \cup_{m\in M} T_pM^*$; it too has a projection map $\pi^* : T^*(M) \to M$ defined by $\pi^*(\omega) = p$ if $\omega \in T_pM^*$. The tangent and cotangent bundles have natural differentiable structures induced by the differentiable structure on M.

A brief exposition of the details follows.[29] If (U, ξ) is a coordinate system on M, then $\phi : \pi^{-1}(U) \to \mathbb{R}^{2n}$, given by

$$\phi(v) = (x^1(\pi(v)), \ldots, x^n(\pi(v)), dx^1(v), \ldots, dx^n(v)),$$

and $\phi^* : \pi^{*-1}(U) \to \mathbb{R}^{2n}$, given by

$$\phi^*(\omega) = (x^1(\pi^*(\omega)), \ldots, x^n(\pi^*(\omega)), \omega(\partial_1), \ldots, \omega(\partial_n)),$$

let us define the topology on $T(M)$ to be that generated by the basis

$$\{\phi^{-1}(U) | \phi \leftrightarrow \xi, U \text{ is open in } \mathbb{R}^{2n}\},$$

and likewise for $T^*(M)$. Under these topologies, $T(M)$ and $T^*(M)$ are second-countable, $2n$-dimensional locally Euclidean spaces. The ϕs become local homeomorphisms and, in fact, a set of atlases; the differentiable structure on $T(M)$ is the maximal atlas containing the ϕs.[30] Similar constructions apply to the so-called *tensor bundle (of type* (r,s) *over* M $T^r_s(M)$, *exterior* k *bundle over* M $\Lambda_k(M)$, and *exterior algebra bundle over* M $\Lambda(M) = \cup_k \Lambda_k(M)$.

The definition of smoothness given above can be checked to be consistent with the definition given at the end of Section 2.3.

We would like to generalize these ideas. To that end, we shall – as is usual – abstract the essential properties of tensor fields and tangent spaces, axiomatize them, and see where they shall lead us. In this particular case, the tangent space and the tensor spaces it induces are vector spaces attached to the manifold at each point. So let us consider an arbitrary vector space at each point of the manifold along some reasonable properties, such as a smooth projection and local triviality.

7.2. **Definitions.** Set M an n-dimensional manifold and V a k-dimensional vector space, and let E be a manifold and π a C^∞ map $E \to M$. On the map π we place two conditions:

- for each $p \in M$, $\pi^{-1}(p)$ is a subset of E diffeomorphic to V when V is considered a manifold;
- for each p there is a neighborhood U with diffeomorphism $\phi : \pi^{-1}(U) \to U \times V$ such that, if $q \in U$, $v \to \phi(q, v)$ is a vector space isomorphism from V to $\pi^{-1}(q)$.

These spaces and conditions together give us a k-vector bundle over M. In shorthand, we may say $\pi : E \to M$ is a k-vector bundle, or simply that E is a k-vector bundle.

We call M the *base manifold*, E the *total manifold*, π the *projection*, $\pi^{-1}(p)$ the *fiber over* p, V the *standard fiber*, and ϕ a *bundle chart*. (Think of the bundle charts ϕ as analogous to coordinate patches from $\mathbb{R}^n$ to M: they let us use properties of $U \times V$ without requiring that E has those properties everywhere, much like coordinate patches permit us to use properties of $\mathbb{R}^n$ without requiring that M be diffeomorphic to $\mathbb{R}^n$.)

[29] Here, I follow Warner's construction (p. 19).

[30] In a sense, we pulled the topology from Euclidean space up through M via the projection map.

If there are two trivializations on overlapping neighborhoods U_A and U_B, are they consistent? Let[31] $\phi_A : \pi^{-1}(U_A) \to U_A \times V$ and $\phi_B : \pi^{-1}(U_B) \to U_B \times V$. The composition on its domain of definition is $\phi_A \circ \phi_B^{-1} : (U_A \cap U_B) \times V \to (U_A \cap U_B) \times V$ can be written as $(\phi_A \circ \phi_B^{-1})(p, v) = (p, g_{AB}(v))$, where g_{AB} is an automorphism of V. In order to maintain consistency, for any diffeomorphism of this type, we must have $g_{AB} \circ g_{BC} = g_{AC}$, which is termed the *cocycle condition.* (The group of automorphisms between local trivializations is called the structure group.)

Two vector bundles over an open set M are isomorphic if there is a map between them which commutes (via their projections) with the identity map on M.

A *section* of E is a map $X : U \to E$ with $\pi \circ X = id_M$. We denote the set of all sections of E over M by $\Gamma(E)$. A set of k linearly independent local sections over U forms a *local basis* or *local frame field* for E. Note that $\phi : U \times V \to \pi^{-1}(U)$ corresponds to exactly one basis for E over U, namely $\phi(p, v) = \sum_i v_i s_i(p)$ (at p, express each point v of $\pi^{-1}(p)$ in terms of the basis $s_i(p)$). Vice-versa holds as well: if there is a trivialization over U, that implies a choice of basis on each fiber over U. If there is a *global* frame field (i.e., one with $U = M$), then we say that E is the *trivial bundle* $M \times V$. (Note that this is not generally the case!)

We can rephrase tensor differentiation in terms of a vector bundle (in this case, the graded tensor bundle over M). We define a *connection in* E to be a bilinear map $\nabla : \mathfrak{X}(M) \times \Gamma(E) \to \Gamma(E)$ such that $\nabla(fX, s) = f\nabla(X, s)$ and $\nabla(X, fs) = f\nabla(X, s) + (Xf)s$ for all $X \in \mathfrak{X}(M)$ and $s \in \Gamma(E)$. In the trivial bundle $M \times V$, it is not difficult to discover a connection. If $(s_1, \ldots, s_n)$ is a frame field and each $s_i(p) = \partial_i|_p$, define $\nabla_X s_i = 0$ for each $X \in \mathfrak{X}(M)$. Then every section $s = \sum a_i s_i$, and when X is given define $\nabla_X s = \sum (Xa_i)s_i$. (This is exactly the covariant derivative we discussed above in motivating the Levi-Civita connection; in this context, it is called the *trivial connection.*)

We have *assumed* that E is locally trivial - namely, at each point of M there is a neighborhood U with a lifting into E that is diffeomorphic to $U \times V$. *On each such neighborhood* U_a, covering M and indexing $a \in A$ by some index set, we take ∇^a to be the trivial connection in each $\pi^{-1}(U_a)$. If $\{f_a\}$ is a partition of unity for the cover $\{U_a\}$, define

$$\nabla_X s = \sum_a f_a \nabla_X^a s.$$

Using a connection, we can define curvature tensors as above. For the trivial connection, $\nabla_X \nabla_Y - \nabla_Y \nabla_X = \nabla_{[X,Y]}$. In general, this doesn't hold, so we measure the departure of E from triviality (under ∇) by

$$R(X, Y) = \frac{1}{2}\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]}.$$

The symmetries from Section 5.3 hold. Note that this result is more general: we make use of neither the metric nor the Levi-Civita connection.

7.3. **The structure equations.** If ∇ is a connection in E and R is its curvature, we can represent them locally by differential forms. Over an open $U \subseteq M$, take a frame field $s_1, \ldots, s_n$. For any vector field X,

$$\nabla_X s_j = \sum_{i=1}^{n} \omega_j^i(X) s_i.$$

[31]With diffeomorphisms it doesn't matter which way you go.

(This makes sense: the covariant derivative of a vector field – in this case the frame field s_j – is again a vector field, so we can of course write it in terms of the frame field.) Since connections are $\mathcal{F}$-linear in the first argument (X), we have that $\omega^i_j(fX) = f\omega^i_j(X)$, and so ω^i_j is a one-form for each i and j. We gather them into the matrix $\omega = (\omega^i_j)$, and call it the *connection form of* ∇ *on* U. Morita suggests we think of it as a one-form on U with values in $\mathfrak{gl}(n, \mathbb{R})$.

We can do the same thing with the curvature.[32] Define Ω^i_j by $R(X,Y)(s_j) = \sum_i \Omega^i_j(X,Y)s_i$. Since Ω is skew-symmetric and $\mathcal{F}$-linear in both its arguments, it is a two-form on U. And, again, we can write it as a matrix: $\Omega = (\Omega^i_j)$.

There is a deep and subtle relationship between the two "form matrices": the *structure equation*. For a vector bundle E, the following equation holds:

$$d\omega = -\omega \wedge \omega + \Omega.$$

(Componentwise,

$$d\omega^i_j = -\sum_{k=1}^{n} \omega^i_k \wedge \omega^k_j + \Omega^i_j.)$$

If ∇ is a connection in a vector bundle E, and there are two subsets U_A and U_B, recall that we can express a diffeomorphism between $\pi^{-1}(U_A)$ and $\pi^{-1}(U_B)$ by $\phi(p,v) = (p, g_{AB}v)$ for some operator $g_{AB} : U_A \cap U_B \to Gl(n, \mathbb{R})$. The frame fields on U_A and U_B each induce connection and curvature forms $\omega_A, \omega_B, \Omega_A, \Omega_B$. We can transform between these forms by

$$\omega_B = g^{-1}_{AB}\omega_A g_{AB} + g^{-1}_{AB} dg_{AB}$$

and

$$\Omega_B = g^{-1}_{AB}\Omega_A g_{AB}.$$

The proof can be found in [6], p. 189. The structure equations — first developed by Cartan — unify curvature and the connection in terms of differential forms in a simple and easily-understood expression.

Part 2. The Physics

8. Introduction

Many physical phenomena can be modeled with the techniques of differential geometry. We begin with electromagnetism, use it to motivate the development of special and general relativity, and conclude with a brief survey of gauge theories, the modern geometric techniques used to model the fundamental forces of the universe.

9. Electricity & Magnetism

We develop the theory of electromagnetism in two ways: first, as an empirical statement, Maxwell's equations, which we then express in forms; second, as an afterthought of $U(1)$ gauge invariance in nonrelativistic quantum mechanics.

[32] We are not thinking here of R as a $(1,3)$ tensor, but instead as an operator which takes three vector fields – X, Y, s_j – to another vector field – $\sum \Omega^i_j s_i$.

9.1. **Maxwell's equations.** Electromagnetics is concerned with the behavior of vector fields in space. In particular, it deals with the relationship between the electric field E, the magnetic field B, the charge density ρ, and the current density J. For a readable development of the classical theory, see [11]; this section loosely follows that reference, Ch. 10.1.

Maxwell's equations are the *empirical facts* of electromagnetism, phrased here as vector equations:

$$\begin{cases} \nabla \cdot E = \frac{\rho}{\epsilon_0} \\ \nabla \times E = -\frac{\partial B}{\partial t} \\ \nabla \cdot B = 0 \\ \nabla \times B = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \end{cases}$$

Writing the metric equivalents of E and B as, respectively, $\mathcal{E}$ and $\mathcal{B}$, we can replace the use of gradients with the Hodge operator $*$ and the exterior derivative:

$$\begin{cases} *d*\mathcal{E} = \frac{\rho}{\epsilon_0} \\ *d\mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t} \\ *d*\mathcal{B} = 0 \\ *d\mathcal{B} = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \end{cases}$$

We can make use of the de Rham cohomology in finding an alternative manner of writing these equations. Since $*d*\mathcal{B} = 0$ and $*$ is an isomorphism, $d*\mathcal{B}$ is zero. That is, $*\mathcal{B}$ is a closed form. In $\mathbb{R}^3$, the second cohomology group is trivial, so there is an exact form $\mathcal{A}$ with $*\mathcal{B} = d\mathcal{A}$, or $\mathcal{B} = *d\mathcal{A}$. This is the one-form metrically equivalent to the *vector potential* A, which has $\nabla \times A = B$.

Similarly, because $*d\mathcal{E} = -\frac{\partial \mathcal{B}}{\partial t} = -(*d\frac{\partial \mathcal{A}}{\partial t})$, we have $*d(\mathcal{E} + \frac{\partial \mathcal{A}}{\partial t}) = 0$. Again, because $*$ is an isomorphism, $\mathcal{E} + \frac{\partial \mathcal{A}}{\partial t}$ is a closed one-form; since the first homology group is also trivial, there is a function V with $\mathcal{E} + \frac{\partial \mathcal{A}}{\partial t} = -dV$. Hence $\mathcal{E} = -(dV + \frac{\partial \mathcal{A}}{\partial t})$.

We may therefore rewrite Maxwell's equations as

$$\begin{cases} *d*dV + \frac{\partial}{\partial t}(*d*A) = -\frac{\rho}{\epsilon_0} \\ \left(*d*d - \mu_0\epsilon_0 \frac{\partial^2}{\partial t^2}\right)\mathcal{A} - d\left(*d*\mathcal{A} + \mu_0\epsilon_0\frac{\partial V}{\partial t}\right) = -\mu_0 J. \end{cases}$$

Once the existence of at least one potential is established, there is some flexibility in the *choice* of potential because the sum of a closed form and an exact form is again closed. Therefore, we may replace $\mathcal{A}$ by $\mathcal{A}' = \mathcal{A} + \alpha$ and V by $V' = V + \beta$ as long as α and β are exact. Again, making use of the fact that the first homology group is trivial, $\alpha = df$ for some $f \in \mathcal{F}$. Because $\mathcal{E} = -(dV + \frac{\partial \mathcal{A}}{\partial t}) = -(dV' + \frac{\partial \mathcal{A}'}{\partial t})$, we have

$$d\left(\beta + \frac{\partial f}{\partial t}\right) = 0.$$

Hence the form in the parentheses is a closed function, independent of position; it futher permits us to write $\beta = -\frac{\partial f}{\partial t} + k(t)$ for some function of time k. Absorbing the function into f, the permissible potentials are parametrized by $\mathcal{F}$:

$$\mathcal{A}' = \mathcal{A} + df \text{ and } V' = V - \frac{\partial f}{\partial t}.$$

This flexibility in the choice of potential is known as *gauge invariance*.

In a vacuum,[33] $\rho = J = 0$ and Maxwell's equations reduce to

$$\begin{cases} \nabla \cdot E = \nabla \cdot B = 0 \\ \nabla \times E = -\frac{\partial B}{\partial t} \\ \nabla \times B = \mu_0 \epsilon_o \frac{\partial E}{\partial t}. \end{cases}$$

By judiciously applying vector calculus identities, we arrive at two identical equations for E and B:

$$\begin{cases} \nabla^2 E = \mu_0 \epsilon_0 \frac{\partial^2 E}{\partial t^2} \\ \nabla^2 B = \mu_0 \epsilon_0 \frac{\partial^2 B}{\partial t^2} \end{cases}$$

This is the wave equation in each component of E and B, which implies that vacuum solutions of Maxwell's equations are waves with speed

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}.$$

This poses an interesting question: if μ_0 and ϵ_0 are fundamental constants of nature — discovered by measuring the relationship between field strength and distance from source — then we expect that, like the gravitational constant G, they will not vary between inertial frames. But this implies that *all* inertial observers must agree on the speed of electromagnetic waves, *regardless* of relative motion. Somehow, the speed of electromagnetic disturbances is a fundamental property of space, not related to the properties of the propagation medium like other waves, and hence not subject to rational disagreement between differently moving observers. As a result — which we will explore when we come to relativity — space and time themselves conspire to ensure the speed of electromagnetic disturbances remains constant for all inertial observers.

9.2. **Electromagnetism and non-relativistic quantum mechanics.** This interesting argument is related in [12], pp. 41-3.

Suppose we know Schrödinger's equation for a free particle:

$$\frac{-\hbar^2}{2m} \nabla^2 \psi = i\hbar \frac{\partial \psi}{\partial t}.$$

The wavefunction ψ lives in L^2 Hilbert space; observables, that is, quantities that can be measured, are expectation values of the wavefunction: $\langle f \rangle = \int \psi^* f \psi$. These are obviously invariant under the rotation $\psi \mapsto e^{i\theta}\psi$ for all real θ: variation in the phase angle of a wavefunction has no physical meaning, much as variation in potential energy by a constant has no physical meaning since the only measurable aspect of potential energy is its change over space or time.

The physical insignificance of global phase invariance leads us to ask: is *local* phase invariance physically insignificant? It would not seem to change expectation values of observables. What happens if we require that θ is a function of position[34] so $\psi(p) \mapsto e^{i\theta(p)}\psi(p)$?

Differentiation causes Schrödinger's equation to pick up extra terms:

$$\frac{-\hbar^2}{2m} \nabla^2 (e^{i\theta}\psi) = i\hbar \frac{\partial e^{i\theta}\psi}{\partial t},$$

[33] This portion of the section is inspired by [11], Ch. 9.1.

[34] Spatial, not temporal – this is *non*-relativistic.

so

$$\frac{-\hbar^2}{2m}e^{i\theta}\nabla^2\psi - \frac{\hbar^2}{2m}e^{i\theta}\sum_j\left(\psi\frac{\partial^2\theta}{\partial x_j^2} + \psi\left(\frac{\partial\theta}{\partial x_j}\right)^2 + 2i\frac{\partial\theta}{\partial x_j}\frac{\partial\psi}{\partial x_j}\right) = i\hbar e^{i\theta}\frac{\partial\psi}{\partial t}.$$

That ugly, *ugly* extra term is the result of θ's ability to vary from point to point. So we replace the normal gradient ∇ by the operator $D = \nabla + ieA$, where e is not the exponential but the charge of the wavefunction and A is some[35] vector field that transforms thus: $A \mapsto A - (1/e)\nabla\theta$. By this token, under the phase change the *operator* changes: $D = \nabla + ieA \mapsto \nabla + iA - i\nabla\theta$.

What does this new operator do to the local phase transformation? A quick check shows that $D(\psi) \mapsto D(e^{i\theta}\psi) = \nabla(e^{i\theta}\psi) + (ieA - i\nabla\theta)\psi = ie^{i\theta}\psi\nabla\theta + e^{i\theta}\nabla\psi + e^{i\theta}ieA\psi - ie^{i\theta}\psi\nabla\theta = e^{i\theta}(\nabla + ieA)\psi = e^{i\theta}D\psi$.

By this token, replacing the gradient operator with D, we have a Schrödinger equation *invariant* under local phase transformations:

$$\frac{-\hbar^2}{2m}D^2\psi = i\hbar\frac{\partial\psi}{\partial t}.$$

We have preserved the *observational invariance* under local phase transformations, but at a cost: we have been forced to critically alter Schrödinger's equation. Differentiation, which canonically representes momentum, has acquired a new term, which indicates that *momentum* has changed. Physically, a free particle is a particle entirely without outside influence; it is impossible to distinguish its wavefunction at one point from another point. In order to do so, we must introduce an outside influence, to separate some locations from others; it manifests itself as the potential A. Now the particle is no longer free; it may be *unbound*, but it is forever coupled to A.

As noted, A is interpreted as the electromagnetic vector potential; hence, the inclusion of A in the invariant Schrödinger equation suggests that the electromagnetic field is somehow coupled to the wavefunction ψ. In fact, we can break the operator D down to its constituents and interpret A as an *actual* potential:

$$\frac{-\hbar^2}{2m}\left|\nabla\psi - eA\right|^2\psi = i\hbar\frac{\partial\psi}{\partial t},$$

which nearly fits the form of the Hamiltonian operator with the potential $V = -2eA + (e|A|)^2$. The Aharonov-Bohm effect[36] documents the *reality* of the vector *potential*: in contradiction to the classical interpretation of the vector potential A as a mathematical convenience, A actually possesses physical significance in quantum mechanics.

10. Special Relativity

This description of special relativity follows [14], Ch. 1, 2.

As described in Section 9.1, vacuum disturbances of electric and magnetic fields propagate at a constant speed for all observers. This poses a problem for transformations between coordinate systems that are moving at a constant speed v relative to each other. Galilean transformations, which correct for the spatial influence of relative motion, but assume that

[35] By "some", I mean the electromagnetic vector potential.

[36] [13] explores the significance of A in problems 4.59-61 and section 10.2.3.

time remains the same for all observers, are no longer sufficient: for under a Galilean transformation, velocities add, so if one observer notes an electromagnetic disturbance propagating at c, a second observer moving at v relative to the first in the same direction as the electromagnetic disturbance should observe the disturbance moving at speed $c+v$, in contradiction to his measurements of μ_0 and ϵ_0. The fact of electromagnetism and the use of Galilean transformations cannot be reconciled.

To illustrate this point with a concrete example, let us work out the wave equation in two coordinate systems under a Galilean transformation.

Let $\phi = \{t, x, y, z\}$ be a coordinate system. The wave equation declares that ϕ observes the electric field E to satisfy the relationship $E_{tt} = E_{xx} + E_{yy} + E_{zz}$. Let us simplify it so we only work with two coordinates[37], $E_{tt} = E_{zz}$, by assuming that E depends only on the z coordinate. We wish to find the form this relationship takes in a coordinate frame $\psi = \{\tau, \xi, \nu, \zeta\}$ which is related to ϕ by the transformation

$$\begin{pmatrix} \tau \\ \zeta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ vt & 1 \end{pmatrix} \begin{pmatrix} t \\ z \end{pmatrix}.$$

In other words, ψ is traveling up on ϕ's z-axis. Let us express E in terms of $\mathcal{E}$, the electric field that ψ sees, $\mathcal{E}(\tau, \zeta) = E(\tau, \zeta + vt) = E(t, z)$, and differentiate accordingly:

$$E_{tt} = \mathcal{E}_{\zeta\zeta};$$

$$E_{zz} = \mathcal{E}_{\tau\tau} - 2v\mathcal{E}_{\tau\zeta} + v^2\mathcal{E}_{\zeta\zeta}.$$

Therefore, in the moving observer's frame, the wave equation takes the form

$$\mathcal{E}_{\zeta\zeta} = \mathcal{E}_{\tau\tau} - 2v\mathcal{E}_{\tau\zeta} + v^2\mathcal{E}_{\zeta\zeta}.$$

The two frames are observing *the very same electric field* with *the very same oscillation.* Since neither observer believes that he is in motion, we have two different versions of the same phenomenon. This is physically unacceptable.

In order to solve this paradox, Einstein he reaffirmed that all observers must formulate physical laws in the same manner *and*, as a consequence, he asserted that Maxwell's laws must be the *same* in all coordinate systems. In particular, the speed of light must be the same for all observers.

This entirely changes the form of coordinate system transformations. If we wish to preserve c under a coordinate transformation, we must have a transformation matrix that preserves the worldline of any photon. In other words, let two observers G and R agree that G omits a photon at a particular moment, and both observers set the origins of their coordinate systems at the event of G releasing a photon. Then their coordinate systems *must agree* on the trajectory the photon travels.

In other words, there exists a transformation B between their coordinate systems such that the worldline of the photon is an eigenspace and the velocity of a photon is preserved (i.e., is an eigenvalue). These are the *Lorentz transformations.*

[37] Lazy! Or are we just being clever?

10.1. **Derivation of two-dimensional Lorentz transformations.** Let A and B be traveling through $\mathbb{R}^2$ at some relative velocity $v < c$. There is a linear transformation[38] between the two (natural Euclidean) coordinate systems A and B have chosen. Let us denote it T; it maps *from* B's system *to* A's system.[39]

Fix A's coordinate system as (t, z) and B's as (τ, ζ). The linear transformation is

$$T = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \tau \\ \zeta \end{pmatrix} = \begin{pmatrix} t \\ z \end{pmatrix}.$$

There are three conditions we impose on the transformation in addition to linearity. First, the slope of B's worldline in A's system must be v. Second, the light cone is preserved – that is, the vectors[40]

$$U = \begin{pmatrix} 1 \\ c \end{pmatrix} \text{ and } D = \begin{pmatrix} 1 \\ -c \end{pmatrix}$$

are eigenvectors of T. Third,[41] $T^{-1} = FTF$, where $F : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$F \begin{pmatrix} z \\ t \end{pmatrix} = \begin{pmatrix} -z \\ t \end{pmatrix}.$$

(Note that $F = F^{-1}$.)

Applying the first condition, we see that

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \tau \\ 0 \end{pmatrix} = \begin{pmatrix} \tau\alpha \\ \tau\beta \end{pmatrix} = \begin{pmatrix} t \\ z \end{pmatrix}$$

so $v = \beta/\alpha$. In particular[42], $\alpha \neq 0$.

Applying the second condition and assuming that associated to U and D are eigenvalues λ_U and λ_D respectively, we arrive at the four linear equations

$$\begin{aligned} \alpha + c\beta &= \lambda_U \\ \gamma + c\delta &= c\lambda_U \\ \alpha - c\beta &= \lambda_D \\ \alpha - c\beta &= -c\lambda_D. \end{aligned}$$

(Note that we have six unknowns and four variables, which is why we need the two conditions on symmetry to exactly specify the system in terms of v. The relative velocity is itself an

[38]Linearity is justified on physical grounds: the laws of nature have to be the same for *both* observers, so they can't change, e.g., from first- to second-order.

[39]We go from B to A because we suppose that *we* are A, and we wish to compare observations B has made to those we have made.

[40]U for "up" and D for "down".

[41]This requires that space be symmetric about the time axis: if both observers stand on their heads, turn around, or whatever, the universe remain the same.

[42]If $\alpha = 0$, then $det(T) = 0$.

unknown, but it is an assumed unknown.) Solving the linear system, we have:

$$\alpha = \frac{\lambda_U + \lambda_D}{2}$$
$$\gamma = c\frac{\lambda_U - \lambda_D}{2}$$
$$\beta = \frac{1}{c}\frac{\lambda_U + \lambda_D}{2}$$
$$\delta = \frac{\lambda_U + \lambda_D}{2},$$

which reduces to $\alpha = \delta$ and $\gamma = c^2\beta$. Hence, applying the first condition,

$$T = \begin{pmatrix} \alpha & \gamma/c^2 \\ \gamma & \alpha \end{pmatrix} = \begin{pmatrix} \alpha & \alpha v/c^2 \\ \alpha v & \alpha \end{pmatrix} = \alpha \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}.$$

Finally, we apply condition three. $T^{-1} = FTF$, or $FT^{-1}F = T$, so $det(T) = det(FT^{-1}F) = det(F)det(T^{-1})det(F) = det(T^{-1})$ since $det(F) = 1$. Therefore, $det(T) = 1$, so

$$\begin{aligned} 1 = det(T) &= det\left(\alpha \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}\right) \\ &= \alpha^2 det \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \\ &= \alpha^2(1 - v^2/c^2). \end{aligned}$$

This finally gives us $\alpha = [\sqrt{1-(v^2/c^2)}]^{-1}$, so

$$T = \frac{1}{\sqrt{1-(v^2/c^2)}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}.$$

We often denote

$$\gamma = \frac{1}{\sqrt{1-(v^2/c^2)}}.$$

The preceding discussion can be found in [14], p. 38.

10.2. **The Minkowski metric and hyperbolic transformations.** The assertion that the Lorentz group of transformations ($SO(1,3)$) on $\mathbb{R}^4$ are the isometries of space and time is equivalent to asserting that $\mathbb{R}^4$ carries the Minkowski metric

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

In other words, if the Lorentz group preserves a metric, that metric *is* the Minkowski metric; if the Minkowski metric is imposed, the Lorentz group (by definition, in some formulations of special relativity) is the isometry group of $\mathbb{R}^4$.

At this point, we are still considering $\mathbb{R}^4$ as a vector space. In the absence of matter, we will often consider $\mathbb{R}^4$ to be the 4-*manifold* M endowed with the Minkowski metric on each tangent plane. In this case, M is isometric to its tangent plane at each point.

10.3. **Kinematic and dynamic consequences.** Following [14], pp. 73-82, in an exercise familiar to any physics undergraduate, we develop the kinematic and dynamical consequences of special relativity by calculating how kinematic and dynamical quantities transform under the Lorentz group.[43]

The moral of this story is twofold: physically, once *all* observers agree on a particular quantity, the dimensions of nature itself conspire to ensure that the quantity is the same for all observers. Mathematically, it is simply a matter of transforming the underlying space and time measurements from which the physical quantity is derived.

Let us, as in the simplified derivation of the Lorentz transformations above, be in coordinate frame A and our hypothetical moving observer have coordinate frame B.

Let E be a spacelike event (i.e., closer to the space spanned by the space axes than the light cone, or one that cannot be reached without traveling faster than c). Then there is some observer who says that E occurs simultaneously with the origin O. This observer is simply the observer whose space axes run through E: if we assume E is in the (z, t) plane and B has aligned his coordinate axes with ours, because E is spacelike, it can be reached by a hyperbolic rotation of some angle.

If an event occurs on the image of B's time axis at some time τ, then we in A observe it occuring at the time t given by the simple transformation

$$\begin{pmatrix} t \\ z \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \begin{pmatrix} \tau \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma\tau \\ v\tau \end{pmatrix}.$$

In other words, assuming clocks are initially synchronized (i.e., coordinate systems share an origin), what B observes to occur in a unit of time, A observes to occur in the fraction γ of a unit of time.

Similarly, let B be holding a ruler along the direction of travel. Its length, as B observes it, is some distance ζ; if one end is at the origin, its other end corresponds to some event $(0, \zeta)$ in B's frame. In A's frame, since one end is at the origin and it is held along the direction of motion, it also lies along A's z-axis at some distance l from the origin. We must have that[44]

$$\begin{pmatrix} \tau \\ \zeta \end{pmatrix} = \gamma \begin{pmatrix} 1 & -v/c^2 \\ -v & 1 \end{pmatrix} \begin{pmatrix} 0 \\ l \end{pmatrix} = \begin{pmatrix} -\gamma l v/c^2 \\ \gamma l \end{pmatrix},$$

or $l = \zeta/\gamma$. So A sees B's ruler *contracted* by a factor of γ: the Lorentz contraction.

Forces present some difficulty in the relativistic scheme. In our rest frame, the equation $f = dp/dt$ holds; morally, because of the principle of relativity, we believe that it must also hold in all other inertial frames. How can we formulate Newton's laws *covariantly* so that they transform relativistically?

For the answer, we turn to the laws of conservation of momentum and energy. Since they hold in an inertial frame, we require them to hold in every inertial frame. Start by identifying the "4-velocity" of a particle moving at constant velocity through A's reference frame as $v = (c, v_x, v_y, v_z)$. The integral curve is the path of the particle; it happens to be a straight line since we are dealing with inertial systems. Let the particle have mass m according to A, and do *not* assume that the particle's mass is independent of relative velocity.

[43]Another method proceeds by assuming that a particle's Lagrangian is relativistically invariant.

[44]The matrix, the transformation from A to B, is also the inverse of the Lorentz transformation from B to A, as the reader may easily verify.

Define the "4-momentum" of the particle by $p = (mc, mv_x, mv_y, mv_z)$. For simplicity's sake, assume velocity is all in the z direction of A's frame; then we can consider $p = (mc, mv)$.

Let us label the rest frame of the particle B. In B, the particle has mass μ and zero velocity, so in B its 4-momentum is $p = (\mu, 0)$. We need this to transform:

$$\begin{pmatrix} m \\ mv \end{pmatrix} = \gamma \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma\mu \\ \gamma\mu v \end{pmatrix}.$$

Among other things, this implies so-called mass dilation: $m = \gamma\mu$. [15], p. 90, notes that the distinction between relativistic mass and rest mass has fallen by the wayside in favor of a deeper focus on relativistic momentum. Defining momentum in terms of the rest mass μ, we have that $p = \gamma(\mu c, \mu v)$ where v may be taken as a vector if appropriate.

Now conservation of momentum *within* a given inertial frame follows as a consequence of the linearity of the coordinate transformation, even though total momentum may have different values in different frames.

Conservation of energy is subsumed into conservation of momentum: we think of energy as a measure of momentum in the *time* dimension – $p = (E/c, \mu v)$.

10.4. **Relativistic electromagnetism in Minkowski space.** Now that we have developed some simple consequences of special relativity, following [11], Ch. 12.3, and [1], Ch. 4, we wish to write Maxwell's equations — which motivated the development of the theory — in a manner conducive to relativistic manipulation. We begin by noting that electric and magnetic fields are not static; they transform into each other under Lorentz boosts. Therefore, a combined tensorial description is in order. We may represent electric and magnetic fields as a 2-form F over Minkowski space. In orthonormal coordinates, the tensor is called F and has elements

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}.$$

Charge and current densities transform exactly like vectors: At a point, ρ is the charge density in the rest system; there is no current. Then in the frame of an observer moving at relative velocity v, the charge density is $\gamma\rho$ and the current density is $J = \gamma\rho v$. Note that ρ itself changes because of length contraction-induced volume changes, but total charge is conserved between frames. We therefore define the "4-current" $J = (c\rho, J_x, J_y, J_z)$. This is in keeping with the interpretation that objects at rest are moving in the time direction with a speed of c.

In these terms, conservation of charge, governed by the *continuity equation* $\nabla \cdot J = -\frac{\partial\rho}{\partial t}$, becomes

$$\sum_\mu \frac{\partial J^\mu}{\partial x_\mu} = 0.$$

In other words, total 4-current is conserved. This reflects conservation of charge. Maxwell's equations can then be written

$$\sum_\nu \frac{\partial F^{\mu\nu}}{\partial x^\mu} = \mu_0 J^\nu, \sum_\nu \frac{\partial G^{\mu\nu}}{\partial x^\nu} = 0,$$

where $G^{\mu\nu}$ is the dual tensor to $F^{\mu\nu}$.

Since B has zero divergence, there is a vector field A such that $B = \nabla \times A$. Similarly, there is a (more familiar) scalar field V such that $E = \nabla V - \partial A/\partial t$. We can create a 4-potential $A^\mu = (V/c, A_x.A_y, A_z)$. Then[45] $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ Now, Maxwell's laws are

$$\partial^\mu(\sum_\nu \partial_\nu A^\nu) - \sum_\nu \partial^\nu(\partial_\nu A^\mu) = \mu_0 J^\mu.$$

If we choose the Lorentz gauge $\sum_\mu \partial_\mu A^\mu = 0$, we have[46]

$$\sum_\nu \partial^\nu \partial_\nu A^\mu = -\mu_0 J^\mu.$$

This is the most succinct way of writing Maxwell's equations; it makes full use of the relationships between space and time first implied by the development of electrodynamics, bringing the physics, in some sense, full circle.

11. General Relativity

In developing general relativity, we follow largely [2], Ch. 12; [14] is extremely helpful, but material is not drawn so carefully from it because it uses cumbersome local coordinates.

11.1. Motivation.

Special relativity is inconsistent with accelerating reference frames. This inconsistency is built into the assumptions of the theory: it compares only observers moving at constant speeds relative to each other. Therefore, it cannot deal with, in particular, gravitational fields. Any generalization of special relativity[47] to deal with gravity must retain special relativity's treatment of light while abandoning Newton's description of gravity as action at a distance.

Newton's law of gravity is a *law*, not a *theory*: Newton proposed a pattern, to be empirically tested, in nature, and confirmed the pattern with his own calculations of the Moon's orbit as well as by predicting Kepler's laws (which had been empirically confirmed through Brahe's exhaustive observations). Newton proposed that, between any two massive objects, $F = Gm_1m_2/r^2$ is an attractive force along their common axis. While Newton's descriptive model remained accurate to within observational uncertainty for over a century and a half, it was philosophically unsatisfactory: the force of gravity worked as an action at a distance.

Special relativity puts an end to action at a distance: the collaboration of space and time to ensure that c remains constant — modeled by the use of the Minkowski metric which partitions all of spacetime into causal and non-causal portions — implies that action at a distance violates causality. Reconciling Newtonian gravity and special relativity, then, will generalize both models.

Einstein's first insight was to consider a frame in free-fall. To a freely-falling observer, space appears to be Minkowski. We can therefore *approximate* space locally by Minkowski space. Einstein's brilliant insight was that, *because* free-fall acceleration in a gravitational field is locally *indistinguishable* from motion in no gravitational field at all, it violates Occam's razor to assume that they are distinct. Much as Einstein did away with the universal ether by

[45] $\partial^\mu = \partial/\partial x_\mu$ and $\partial_\mu = \partial/\partial x^\mu$.

[46] Of all things, this is the Klein-Gordon equation for a massless particle: A^μ is the wavefunction of a photon.

[47] Perhaps to be termed "general relativity"?

pointing out that it is simpler to simply postulate that electromagnetic radiation travels at a constant speed through empty space, he did away with the gravitational force with the *Principle of Covariance*: since frames free-falling in a gravitional field are indistinguishable from inertial frames, we should simply assume that they *are* inertial frames.

11.2. **Construction.** Here is general relativity in a nutshell: *freely falling objects follow geodesics.* The effects of gravity, then, can be deduced from tidal forces: the rate at which geodesics diverge from each other, which is related to the curvature of space. Gravity is thus the curvature of space and the consequent distortion of geodesics from the Euclidean notion of straightness.

Gravity is caused by matter. Matter — momentum — is equivalent to energy, which includes kinetic and rest energy. In special relativity, one can be transformed into another by a judicious coordinate change. We represent matter and energy by a symmetric type-$(0,2)$ tensor, T, denoted the *stress-energy tensor.* There are several empirical constraints on T. For any timelike future-pointing unit vector, if v, w are perpendicular to u, $T(v,w)$ represents the flux of energy-momentum in the direction of v across a surface perpendicular to w, much like the classical stress tensor. The energy density $T(u,u)$ is conserved and nonnegative; $divT = 0$ signifies the conservation of mass-energy.

To express the relationship between curvature and matter, Einstein postulated the simplest relationship: $\Gamma = kT$, where Γ is a tensor that depends on Ric and k is a proportionality constant. Because $div(T) = 0$, we must have $div(\Gamma) = 0$; if Γ is to depend on Ric, since $div(Ric) = -\frac{1}{2}dS$, it is rational to set $\Gamma = Ric - \frac{1}{2}Sg$. The proportionality constant turns out to be a mixture of G, π, and c, giving *Einstein's field equations*:

$$Ric - \frac{1}{2}Sg = \frac{8\pi G}{c^4}T,$$

or, in component form,

$$R_{ij} - \frac{1}{2}Sg_{ij} = \frac{8\pi G}{c^4}T_{ij}.$$

Interestingly, general relativity can be derived as a gauge theory of invariance under all transformations with positive determinant; see, e.g., [1], Ch. 5.

12. Gauge Theories

As we have applied mathematical ideas to sketch some physical models, we have seen that demanding symmetry invariance can lead to very powerful generalizations. The notion of *requiring* symmetry invariance was first formalized by Yang and Mills; they advanced the first true gauge theory. The tools of gauge theories, generalized and refined, have led to the most successful theories of particle physics: the electroweak unification of Glashow, Salam, and Weinberg and the theory of quantum chromodynamics. Following [15] Ch. 11 very closely, we here will briefly outline the ideas underlying gauge theories and their construction.

12.1. **Invariant Lagrangians.** The discussion of classical mechanics often begins with Newton's second law: $F = \frac{dP}{dt}$. From it and the functional form of the force, the time evolution of the system can be determined. Quantities such as kinetic and potential energy and laws such as the conservation of energy are secondary, derived concepts. However, there is no reason not to start with energy and derive controlling equations for the system. One such approach, due to Hamilton, relies on the conservation of energy as its starting point; another, due to Lagrange, begins with the *principle of least action.*

Given a field,[48] its state can be specified as a function of fields and their time derivatives — namely, the state $L(\phi_i, \partial_\mu \phi_i)$. This state can be interpreted as a function which varies pointwise, taking as inputs $\phi_i(p)$ and the partial derivatives of ϕ at p; it can thus be interpreted as a collection of 4-forms[49] L_i. The functional, or group of 4-forms, is known as the *Lagrangian density.* The principle of least action dictates that, for each i, the system will evolve in such a manner as to *minimize* the action $\int L$.

Through variational calculus techniques, it can be shown that a solution minimizing the action must satisfy the *Euler-Lagrange equations*[50]

$$\partial_\mu \left(\frac{\partial L}{\partial(\partial_\mu \phi_i)} \right) = \frac{\partial L}{\partial \phi_i}.$$

Let's look at three examples. First, consider the Klein-Gordon Lagrangian for a spin 0 field of mass m:

$$L = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\left(\frac{mc}{\hbar}\right)^2 \phi^2.$$

There is only one field here, so it is not terribly tedious to use the Euler-Lagrange equations:

$$\frac{\partial L}{\partial(\partial_\mu \phi)} = \partial^\mu \phi, \frac{\partial L}{\partial \phi} = -\left(\frac{mc}{\hbar}\right)^2 \phi,$$

so

$$\partial_\mu \partial^\mu \phi + \left(\frac{mc}{\hbar}\right)^2 \phi = 0.$$

This is the controlling equation for the wavefunction of a single free particle of zero spin and mass m. Actual wavefunctions are solutions to this equation.

Second, consider the Dirac Lagrangian for the wavefunction of a particle of spin $\frac{1}{2}$.

$$L = i(\hbar c)\overline{\psi}\gamma^\mu \partial_\mu \psi - (mc^2)\overline{\psi}\psi.$$

We take *two* field, ψ and $\overline{\psi}$, so after cranking through the Euler-Lagrange equations we have *two* sets of equations:

$$i\gamma^\mu \partial_\mu \psi - \left(\frac{mc}{\hbar}\right)\psi = 0$$
$$i\partial_\mu \overline{\psi}\gamma^\mu + \left(\frac{mc}{\hbar}\right)\overline{\psi} = 0.$$

These are the Dirac equation and its adjoint, respectively, for a particle of spin $\frac{1}{2}$ and mass m. They model relativistic electrons and positrons.

---

[48]The classical version actually starts with a single particle $\sigma(t)$ and defines, for kinetic energy $T = \frac{1}{2}m[\sigma'(t)]^2$ and potential energy $U(p)$, $L = T - U = \frac{1}{2}m[\sigma'(t)]^2 - U(p)$. The same minimization procedure as above is followed, and the σ which solves the differential equations is the future evolution of the system. Of course, if F is a conservative force on the particle, it can be shown that, if $F = -\nabla U$, the solution of Newton's second law minimizes L.

[49]Actually, more generally, highest-dimensional forms.

[50]In this section, and this section alone, we use Einstein's summation notation:

$$A^\mu B_\mu = \sum_\mu A^\mu B_\mu.$$

Third. consider the Proca Lagrangian for a spin-1 (vector) field.

$$L = -\frac{1}{16\pi}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) + \frac{1}{8\pi}\left(\frac{mc}{\hbar}\right)^2 A^\nu A_\nu.$$

After applying the minimization condition, we are left with the model

$$\partial_\mu(\partial^\mu A^\nu - \partial^\nu A^\mu) + \left(\frac{mc}{\hbar}\right)^2 A^\nu = 0.$$

This describes a particle of spin 0 and mass m. If $m = 0$, then, substituting $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$, we have $\partial_\mu F^{\mu\nu} = 0$, Maxwell's equations in the vacuum. The vector potential A^μ is the wavefunction of a photon.

12.2. **Gauge invariance.** Let us examine again the Dirac Lagrangian, in a manner very similar to Section 9.2:

$$L = i(\hbar)\overline{\psi}\gamma^\mu\partial_\mu\psi - (mc^2)\overline{\psi}\psi.$$

As in the example above where the classical electromagnetic potential was derived from the Schrödinger equation, we insist on *local* phase invariance on the action. That is, observations should not depend on the phase of the spinor's wavefunction. We therefore examine the effect of a local phase change $\psi \to e^{i\theta(p)}\psi$ on the Lagrangian: $L \to L - \hbar c(\partial_\mu\theta)\overline{\psi}\gamma^\mu\psi$. Setting $\lambda = -\frac{\hbar c}{q}\theta$, where q now represents the charge, we have that

$$\psi \to e^{-iq\lambda/\hbar c}\psi$$

and, correspondingly,

$$L \to L - (q\overline{\psi}\gamma^\mu\psi)\partial_\mu\psi.$$

We wish to make the Lagrangian *in*variant, so to get rid of the extra term we subtract a term which vanishes upon rotation by λ and cancels out the extra term:

$$L = \left[i(\hbar)\overline{\psi}\gamma^\mu\partial_\mu\psi - (mc^2)\overline{\psi}\psi\right] - (q\overline{\psi}\gamma^\mu\psi)A_\mu$$

where A_μ is a vector function, the *gauge field*, which obeys $A_\mu \to A_\mu + \partial_\mu\lambda$.

Now L is invariant under the local phase rotation, but we have introduced a new term into the equation, and we need a free term for it. It is a vector, so we look to the Proca Lagrangian, setting the mass to zero so that the free term transforms correctly:

$$L = \left[i(\hbar)\overline{\psi}\gamma^\mu\partial_\mu\psi - (mc^2)\overline{\psi}\psi\right] - (q\overline{\psi}\gamma^\mu\psi)A_\mu + \left[-\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu}\right].$$

($F^{\mu\nu}$ is as above.)

We have produced a Lagrangian, invariant under phase change (i.e., under action by $U(1)$) which has not one but *three* terms: the original electron/positron term, the coupling term where the wavefunction of the spinor interacts with the vector gauge field, and the free term of a massless particle of spin 0. We, merely by insisting that the Dirac Lagrangian be invariant under action of $U(1)$, have created electrodynamics.

12.3. **Fiber bundles.** The equations describing the transformation of the gauge field under gauge transformations and the requirement that ordinary differentiation be replaced by gauge-covariant differentiation implies that gauge theories could be more precisely modeled as theories of connections on vector bundles, or, more generally, on *fiber bundles*.[51] A fiber bundle is a generalization of a vector bundle that permits the expression of the action of a particular Lie group to influence the connection. The derivative constructed from a connection is called a *covariant derivative*; expressing the gauge invariance of a wavefunction under a symmetry group requires the use of the covariant derivative in the Lagrangian, which in turn implies the existence of a gauge field, a mapping from a manifold into the symmetry group's Lie algebra.

Conclusion

I hope this thesis has served well. It certainly has helped me fit together and better understand the mathematics of differentiable manifolds and how they relate to physics. This project is hardly *complete*, however: the interested reader[52] will certainly need to pursue these ideas in other sources to see, for example, gauge theories take more precise mathematical shape when expressed in the much richer and more general language of fiber bundles.

References

[1] M. Göckeler and T. Schücker. *Differential geometry, gauge theories, and gravity.* Cambridge Monographs in Mathematical Physics. Cambridge University Press, Cambridge, 1990.

[2] B. ONeill. *Semi-Riemannian geometry with applications to relativity.* Pure and Applied Mathematics. Academic Press, Inc., New York. 1983.

[3] F. Warner. *Foundations of differentiable manifolds and Lie groups.* Graduate Texts in Mathematics. Springer-Verlag New York, Inc., New York, 1983.

[4] S. Kobayashi and K. Nomizu. *Foundations of differential geometry, Volumes I and II.* Classics Library Edition. John Wiley & Sons Inc., New York, 1996.

[5] C. Isham. *Modern differential geometry for physicists.* World Scientific Publishing Co., Singapore. 1989.

[6] S. Morita. *Geometry of differential forms.* Translations of Mathematical Monographs, Vol. 201. American Mathematical Society, Providence, RI, 2001.

[7] J. Fraleigh. *A first course in abstract algebra.* Seventh Edition. Pearson Education, Inc., 2003.

[8] B. ONeill. *Elementary differential geometry.* Revised Second Edition. Elsevier Academic Press, Burlington, MA, 2006.

[9] Rudin, Walter. *Principles of mathematical analysis.* Third edition. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Dsseldorf, 1976.

[10] B. Hall. *Lie groups, Lie algebras, and representations: an elementary introduction.* Springer Science+Business Media, 2003.

[11] D. Griffiths. *Introduction to electrodynamics.* Third Edition. Prentice-Hall, Inc., Upper Saddle River, NJ, 1999.

[12] C. Quigg. *Gauge theories of the strong, weak, and electromagnetic interactions.* Frontiers in Physics. The Benjamin-Cummings Publishing Co., Inc., Reading, MA, 1983.

[13] D. Griffiths. *Introduction to quantum mechanics.* Second Edition, International Edition. Pearson Prentice Hall Inc., New Delhi, 2005.

[14] J. Callahan. *The geometry of spacetime: an introduction to special and general relativity.* Springer-Verlag New York, Inc., New York, 2006.

[15] D. Griffiths. *Introduction to elementary particles.* Harper & Row Publishers, Inc., New York, 1987.

[51] [1] develops fiber bundles precisely for such applications.

[52] You!